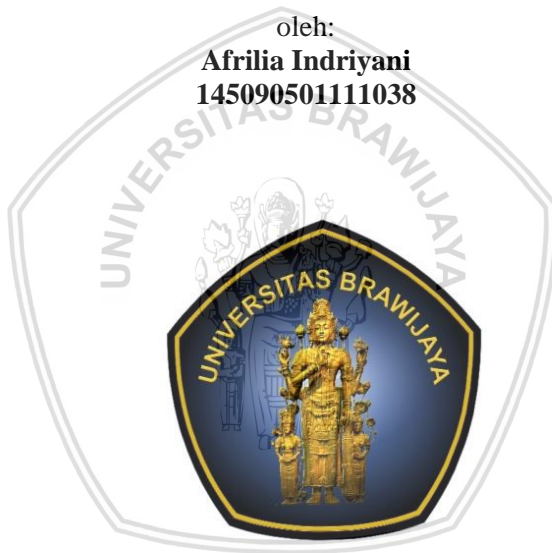


HALAMAN JUDUL**PENENTUAN BANYAKNYA KLASSTER OPTIMAL METODE
AVERAGE LINKAGE MENGGUNAKAN GAP STATISTIC,
SILHOUETTE, DAVIES-BOULDIN****SKRIPSI**

Sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

oleh:

Afrilia Indriyani
145090501111038



**PROGRAM STUDI SARJANA STATISTIKA
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2018**



LEMBAR PENGESAHAN SKRIPSI**PENENTUAN BANYAKNYA KLASSTER OPTIMAL METODE
AVERAGE LINKAGE MENGGUNAKAN GAP STATISTIC,
SILHOUETTE, DAVIES-BOULDIN**

oleh:

AFRILIA INDRIYANI

145090501111038

**Setelah dipertahankan di depan Majelis Penguji pada tanggal 12
Juli 2018 dan dinyatakan memenuhi syarat untuk memperoleh
gelar Sarjana Statistika**

Dosen Pembimbing

Dr. Umu Sa'adah, M.Si
NIP. 196807252002122001

Mengetahui,

**Ketua Jurusan Statistika
Fakultas MIPA Universitas Brawijaya**

Rahma Fitriani, S.Si., M.Sc. Ph.D.
NIP. 197603281999032001



LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Afrilia Indriyani

NIM : 145090501111038

Jurusan : Statistika

Skripsi Berjudul :

**PENENTUAN BANYAKNYA KLASSTER OPTIMAL METODE
*AVERAGE LINKAGE MENGGUNAKAN GAP STATISTIC,
SILHOUETTE, DAVIES-BOULDIN***

Dengan ini menyatakan bahwa:

1. Isi dari Skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termasuk di isi dan tertulis di daftar pustaka Skripsi ini.
2. Apabila dikemudian hari ternyata Skripsi yang saya tulis terbukti hasil jiplakan, saya akan bersedia menanggung resiko.

Demikian pernyataan ini saya buat dengan segala kesadaran.

Malang, 12 Juli 2018

Yang menyatakan,

Afrilia Indriyani

NIM. 145090501111038



PENENTUAN BANYAKNYA KLASTER OPTIMAL METODE AVERAGE LINKAGE MENGGUNAKAN GAP STATISTIC, SILHOUETTE, DAVIES-BOULDIN

ABSTRAK

Analisis kluster adalah teknik pengklasteran multivariat yang digunakan untuk mengklasifikasi objek atau individu ke dalam kluster yang sama berdasarkan karakteristiknya. Secara garis besar algoritma kluster dalam mengklasifikasi terdapat dua kategori, yaitu metode hirarki dan nonhirarki. Masalah utama dalam analisis kluster adalah menentukan banyaknya kluster optimal. Beberapa penelitian telah menghasilkan metode penentuan banyaknya kluster optimal menggunakan indeks validitas. Pada penelitian ini bertujuan untuk membandingkan tiga indeks validitas yaitu *gap statistic*, *silhouette*, dan *Davies-Bouldin* berdasarkan nilai *Cluster Tightness Measure* (CTM) terkecil dan melihat stabilitas kluster yang terbentuk menggunakan koefisien *Jaccard*. Terdapat empat data yang digunakan dalam penelitian ini. Berdasarkan hasil penelitian diperoleh kesimpulan bahwa indeks *gap statistic* merupakan indeks validitas terbaik karena menghasilkan nilai CTM terkecil.

Kata Kunci: Analisis Kluster, *Cluster Tightness Measure* (CTM), Indeks Validitas, Koefisien *Jaccard*,



DETERMINING NUMBER OF CLUSTER OPTIMAL METHOD AVERAGE LINKAGE USING GAP STATISTIC, SILHOUETTE, DAVIES-BOULDIN

ABSTRACT

Cluster analysis is a multivariate clustering technique used to classify objects or individuals into the same cluster based on their characteristics. Broadly clustered cluster algorithm there are two categories, namely hierarchy and non hierarchy method. The main problem in cluster analysis is to determine the optimal number of clusters. Several studies have resulted in the optimal cluster determination method using the validity index. This study aims to compare three validity indexes: statistic gap, silhouette, and Davies-Bouldin based on the smallest Cluster Tightness Measure (CTM) value and see the cluster stability formed using Jaccard coefficients. There are four data used in this research. Based on the research results obtained conclusion that the index gap statistic is the best validity index because it produces the smallest CTM value.

Keywords: Cluster Analysis, Cluster Tightness Measure (CTM), Validity index, Jaccard Coefficients.



KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas segala limpahan rahmat, karunia dan hidayahNya karena penulis dapat menyelesaikan tugas akhir ini. Penulis Skripsi ini diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Statistika dengan judul "Penentuan Banyaknya Klaster Optimal Metode *Average Linkage* Menggunakan *Gap Statistic*, *Silhouette*, *Davies-Bouldin*". Dalam menyelesaikan skripsi ini banyak pihak telah membantu dan memberi dukungan kepada penulis, untuk itu penulis ingin menyampaikan rasa terimakasih kepada:

1. Dr. Umu Sa'adah, M.Si. selaku dosen pembimbing Skripsi yang telah meluangkan banyak waktu untuk memberikan bimbingan dalam penulisan Skripsi ini.
2. Samingun Handoyo, S.Si., M.Cs selaku dosen penguji I yang telah membantu menyempurnakan penulisan Skripsi ini.
3. Rahma Fitriani S.Si., M.Sc., Ph.D selaku dosen penguji II yang telah membantu menyempurnakan penulisan Skripsi ini.
4. Keluarga yang selalu mendoakan dan memberikan dukungan.
5. Semua sahabat dan teman-teman statistika 2014 yang telah memberikan semangat dan dukungannya. Semua pihak yang telah membantu dalam penulisan Skripsi ini yang tidak dapat penulis sebutkan satu per satu.

Semoga Skripsi ini bermanfaat dalam rangka menambah wawasan keilmuan bagi pembaca dan bagi penulis sendiri.

Malang, Juli 2018

Penulis



DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI.....	iii
LEMBAR PERNYATAAN.....	v
ABSTRAK.....	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR	xvii
DAFTAR TABEL.....	xix
DAFTAR LAMPIRAN	xxiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA	5
2.1 Analisis Biplot.....	5
2.2 Analisis Kluster	6
2.2.1 Konsep Kemiripan	7
2.2.2 Jarak Euclidean	7
2.2.3 Metode Hirarki	8
2.2.4 Dendogram	9
2.3 Indeks Validitas Kluster	10
2.3.1 <i>Gap Statistic</i>	10
2.3.2 Indeks <i>Silhouette</i>	11
2.3.3 Indeks <i>Davies-Bouldin</i>	11
2.3 <i>Cluster Tightness Measure</i> (CTM)	12
2.4 Struktur Stabilitas Kluster	12
BAB III METODE PENELITIAN	15
3.1 Data dan Sumbernya	15
3.2 Metode yang Digunakan	15
3.3 Diagram Alir	17



BAB IV HASIL DAN PEMBAHASAN	19
4.1 Hasil Analisis Biplot	19
4.2 Statistika Deskriptif.....	22
4.3 Hasil Analisis Klaster	26
4.4 Hasil Analisis Indeks Validitas (IV).....	26
4.4.1 IV Untuk <i>Gap Statistic</i>	26
4.4.2 IV Untuk <i>Silhouette</i>	29
4.4.3 IV Untuk <i>Davies-Bouldin</i>	31
4.5 Hasil Perhitungan <i>Cluster Tightness Measure</i> (CTM)33	
4.5.1 CTM Untuk <i>Gap Statistic</i>	33
4.5.2 CTM Untuk <i>Silhouette</i>	35
4.5.3 CTM Untuk <i>Davies-Bouldin</i>	37
4.5.4 Perbandingan Hasil CTM	40
4.6 Karakteristik Data	40
4.7 Hasil Analisis Struktur Stabilitas Klaster	42
4.8 Interpretasi dan Memprofil Klaster	45
BAB V PENUTUP	49
5.1 Kesimpulan.....	49
5.2 Saran.....	49
DAFTAR PUSTAKA	51
LAMPIRAN	53



DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Jumlah Kluster Optimal.....	17
Gambar 3.2 Diagram Alir Stabilitas Kluster	18
Gambar 4.1 Grafik Analisis Biplot Data 1	19
Gambar 4.2 Grafik Analisis Biplot Data 2	20
Gambar 4.3 Grafik Analisis Biplot Data 3	21
Gambar 4.4 Grafik Analisis Biplot Data 4	22





DAFTAR TABEL

Tabel 4.1	Statistika Deskriptif Data 1	22
Tabel 4.2	Statistika Deskriptif Data 2	23
Tabel 4.3	Statistika Deskriptif Data 3	24
Tabel 4.4	Statistika Deskriptif Data 4	25
Tabel 4.5	Nilai Indeks Validitas <i>Gap Statistic</i> Pada Data 1	27
Tabel 4.6	Nilai Indeks Validitas <i>Gap Statistic</i> Pada Data 2	27
Tabel 4.7	Nilai Indeks Validitas <i>Gap Statistic</i> Pada Data 3	28
Tabel 4.8	Nilai Indeks Validitas <i>Gap Statistic</i> Pada Data 4	28
Tabel 4.9	Nilai Indeks Validitas <i>Silhouette</i> Pada Data 1	29
Tabel 4.10	Nilai Indeks Validitas <i>Silhouette</i> Pada Data 2	29
Tabel 4.11	Nilai Indeks Validitas <i>Silhouette</i> Pada Data 3	30
Tabel 4.12	Nilai Indeks Validitas <i>Silhouette</i> Pada Data 4	30
Tabel 4.13	Nilai Indeks Validitas <i>Davies-Bouldin</i> Pada Data 1	31
Tabel 4.14	Nilai Indeks Validitas <i>Davies-Bouldin</i> Pada Data 2	31
Tabel 4.15	Nilai Indeks Validitas <i>Davies-Bouldin</i> Pada Data 3	32
Tabel 4.16	Nilai Indeks Validitas <i>Davies-Bouldin</i> Pada Data 4	32
Tabel 4.17	Simpangan Baku <i>Gap Statistic</i> Pada Data 1	33
Tabel 4.18	Simpangan Baku <i>Gap Statistic</i> Pada Data 2	33
Tabel 4.19	Simpangan Baku <i>Gap Statistic</i> Pada Data 3	34
Tabel 4.20	Simpangan Baku <i>Gap Statistic</i> Pada Data 4	35
Tabel 4.21	Simpangan Baku <i>Silhouette</i> Pada Data 1	35
Tabel 4.22	Simpangan Baku <i>Silhouette</i> Pada Data 2	36
Tabel 4.23	Simpangan Baku <i>Silhouette</i> Pada Data 3	36
Tabel 4.24	Simpangan Baku <i>Silhouette</i> Pada Data 4	37
Tabel 4.25	Simpangan Baku <i>Davies-Bouldin</i> Pada Data 1	37
Tabel 4.26	Simpangan Baku <i>Davies-Bouldin</i> Pada Data 2	38
Tabel 4.27	Simpangan Baku <i>Davies-Bouldin</i> Pada Data 3	38
Tabel 4.28	Simpangan Baku <i>Davies-Bouldin</i> Pada Data 4	39
Tabel 4.29	Nilai CTM Indeks Validitas <i>Gap Statistic</i> , <i>Silhouette</i> , <i>Davies-Bouldin</i>	40
Tabel 4.30	Struktur Stabilitas Klaster Pada Data 1	42
Tabel 4.31	Struktur Stabilitas Klaster Pada Data 2	43
Tabel 4.32	Struktur Stabilitas Klaster Pada Data 3	43
Tabel 4.33	Struktur Stabilitas Klaster Pada Data 4	44
Tabel 4.34	Nilai Rata-Rata Objek Dalam Klaster Pada Setiap Variabel Data 1	45



Tabel 4.35 Nilai Rata-Rata Objek Dalam Klaster Pada Setiap Variabel Data 2..... 46

Tabel 4.36 Nilai Rata-Rata Objek Dalam Klaster Pada Setiap Variabel Data 3..... 47

Tabel 4.37 Nilai Rata-Rata Objek Dalam Klaster Pada Setiap Variabel Data 44..... 48





DAFTAR LAMPIRAN

Lampiran 1	Data Penyakit Provinsi Jawa Timur 2016	53
Lampiran 2	Data Indikator Makro Ekonomi Provinsi Sulawesi Selatan Tahun 2014.....	55
Lampiran 3	Data Jumlah Rumah Tangga Usaha Pertanian yang Melakukan Pengolahan Hasil Pertanian Menurut Provinsi dan Subsektor 2013.....	56
Lampiran 4	Data Jumlah Rasio Murid dan Guru Tahun 2016	60
Lampiran 5	Syntax <i>Package clustersim</i> Program R Metode <i>Gap Statistic</i>	64
Lampiran 6	Syntax <i>Package clustersim</i> Program R Metode <i>Davies-Bouldin</i>	66
Lampiran 7	Grafik Metode <i>Gap Statistic</i> Data 1.....	67
Lampiran 8	Grafik Metode <i>Gap Statistic</i> Data 2.....	68
Lampiran 9	Grafik Metode <i>Gap Statistic</i> Data 3.....	69
Lampiran 10	Grafik Metode <i>Gap Statistic</i> Data 4.....	70
Lampiran 11	Grafik Metode <i>DaviesBouldin</i> Data 1.....	71
Lampiran 12	Grafik Metode <i>DaviesBouldin</i> Data 2.....	72
Lampiran 13	Grafik Metode <i>DaviesBouldin</i> Data 3.....	73
Lampiran 14	Grafik Metode <i>DaviesBouldin</i> Data 1.....	74
Lampiran 15	Hasil Perhitungan CTM <i>Gap Statistic</i> Data 1	75
Lampiran 16	Hasil Perhitungan CTM <i>Gap Statistic</i> Data 2.....	76
Lampiran 17	Hasil Perhitungan CTM <i>Gap Statistic</i> Data 3.....	77
Lampiran 18	Hasil Perhitungan CTM <i>Gap Statistic</i> Data 4.....	78
Lampiran 19	Hasil Perhitungan CTM <i>Silhouette</i> Data 1.....	79
Lampiran 20	Hasil Perhitungan CTM <i>Silhouette</i> Data 2.....	80
Lampiran 21	Hasil Perhitungan CTM <i>Silhouette</i> Data 3.....	81
Lampiran 22	Hasil Perhitungan CTM <i>Silhouette</i> Data 4.....	82
Lampiran 23	Hasil Perhitungan CTM <i>Davies-Bouldin</i> Data 1..	83
Lampiran 24	Hasil Perhitungan CTM <i>Davies-Bouldin</i> Data 2..	84
Lampiran 25	Hasil Perhitungan CTM <i>Davies-Bouldin</i> Data 3..	85
Lampiran 26	Hasil Perhitungan CTM <i>Davies-Bouldin</i> Data 4..	86

BAB I PENDAHULUAN

1.1 Latar Belakang

Sejak dahulu manusia sudah sering melakukan pengklasteran atau klasifikasi suatu objek yang ada di sekitarnya. Pengklasteran pada masa itu dilakukan berdasarkan ciri fisik, sifat, dan manfaat objek bagi manusia. Misalkan pengklasteran yang dilakukan pada makhluk hidup yaitu kambing, sapi dan kerbau dapat diklasterkan ke dalam hewan berkaki empat, pengklasteran tumbuhan berdasarkan manfaat bagi manusia yaitu sebagai penghasil pangan dan obat-obatan. Terdapat keterbatasan ketika mengklasterkan banyak objek, terlebih jika objek tersebut memiliki karakteristik yang berbeda-beda. Oleh karena itu diperlukan metode untuk membagi objek tersebut ke dalam klaster yang lebih kecil sesuai karakteristiknya.

Analisis klaster adalah suatu teknik pengklasteran multivariat yang digunakan untuk mengklasifikasikan objek ke dalam klaster yang relatif homogen berdasarkan karakteristiknya (Hair dkk., 2010). Secara umum analisis klaster dibagi menjadi dua, yaitu analisis klaster hirarki dan nonhirarki. Metode hirarki adalah metode pengklasteran secara terstruktur dan bertahap berdasarkan pada kemiripan sifat antar objek. Kemiripan sifat tersebut dapat ditentukan dari kedekatan jarak. Analisis klaster hirarki terdiri dari dua metode, yaitu penggabungan (*agglomerative*) dan pemisahan (*divisive*). Metode *agglomerative* yang sering digunakan adalah *single linkage*, *complete linkage*, dan *average linkage*. *Single linkage* adalah menggabungkan objek yang memiliki jarak terdekat atau didasarkan atas jarak minimum. *Complete linkage* adalah menggabungkan objek yang memiliki jarak yang paling jauh. *Average linkage* menggabungkan objek menurut jarak rata-rata antara setiap pasangan objek.

Salah satu masalah dalam analisis klaster adalah menentukan banyaknya klaster optimal. Banyaknya klaster optimal ditentukan dengan melakukan validitas klaster. Validitas klaster merupakan untuk mengevaluasi hasil analisis klaster sehingga dihasilkan klaster yang optimal. Terdapat beberapa indeks validitas klaster yaitu: *Gap Statistic*, *index Silhouette*, dan *index Davies-Bouldin*. *Gap statistic* adalah suatu metode untuk estimasi banyaknya klaster optimal dalam

satu kumpulan data. *Gap statistic* digunakan pada hasil pengklasteran dari beberapa metode pengklasteran, misalnya metode hirarki dan K-means. *Silhouette* adalah suatu metode untuk mengevaluasi validitas kluster dan digunakan untuk memilih banyaknya kluster yang sesuai. *Silhouette* tidak dipengaruhi oleh metode pengelompokan yang digunakan. *Davies-Bouldin* adalah untuk mengevaluasi kluster dan digunakan untuk mendapatkan kluster yang optimal dengan jarak intra-kluster minimum. Ketiga metode tersebut dibandingkan berdasarkan nilai *Cluster Tightness Measure* (CTM).

Validitas sangat penting dalam analisis kluster karena metode pengklasteran menghasilkan kluster untuk kumpulan data yang homogen. Aspek penting dalam validitas kluster adalah stabilitas. Menurut Hennig (2006), stabilitas adalah kluster valid yang bermakna tidak akan menghilang dengan mudah, saat kumpulan data diubah dengan cara yang *non-essential*. Perubahan data asli dengan cara yang *non-essential* dilakukan dengan menggunakan metode *bootstrap*. Skema evaluasi stabilitas kluster menggunakan nilai maksimum koefisien *Jaccard* antara masing-masing kluster asli dengan kluster *bootstrap*. Metode *bootstrap* pertama kali diperkenalkan oleh Efron 1979. Pada dasarnya metode *bootstrap* adalah metode pengambilan sampel secara acak dengan pengembalian sebanyak n elemen dari sampel awal yang berukuran n dan diulang sebanyak B kali.

Fitrisiana (2009) melakukan penelitian perbandingan *gap statistic* dan indeks validitas gabungan dalam menentukan banyaknya kluster optimal pada analisis kluster menggunakan data ekspresi gen. hasil penelitian tersebut menunjukkan pada data 1 *gap statistic* lebih baik digunakan sedangkan untuk data 2 kedua metode tersebut memiliki kemampuan yang sama. Sholihah (2011) juga melakukan penelitian tentang penentuan banyaknya kluster menggunakan beberapa indeks validitas. Tibshirani, dkk (2001) mengusulkan metode *gap statistic* dalam menentukan banyaknya kluster optimal. Pada penelitian kali ini akan membandingkan indeks validitas kluster optimal yang paling baik menggunakan *gap statistic*, *silhouette*, dan *Davies-Boudin*. Ketiga indeks tersebut dibandingkan menggunakan nilai *Cluster Tightness Measure* (CTM) dan melakukan evaluasi banyaknya kluster yang terbentuk menggunakan koefisien *Jaccard*.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana estimasi banyaknya kluster optimal menggunakan indeks validitas *Gap Statistic*, *Silhouette*, *Davies Bouldin* pada metode *average linkage*?
2. Indeks validitas apa yang terbaik dalam menentukan banyaknya kluster optimal?
3. Bagaimana estimasi derajat kestabilan untuk mengevaluasi banyaknya kluster yang terbentuk dengan menggunakan Koefisien *Jaccard*.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah

1. Metode kluster yang digunakan adalah agglomerative hirarki berdasarkan metode *average linkage*.
2. Penentuan indeks validitas terbaik berdasarkan nilai *Cluster Tightness Measure* (CTM).

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini sebagai berikut:

1. Mengetahui hasil estimasi banyaknya kluster optimal menggunakan indeks validitas *Gap Statistic*, *Silhouette*, *Davies Bouldin* pada metode *average linkage*.
2. Mengetahui indeks validitas yang terbaik dalam menentukan banyaknya kluster
3. Mengetahui estimasi derajat kestabilan untuk mengevaluasi banyaknya kluster yang terbentuk dengan menggunakan koefisien *Jaccard*.

1.5 Manfaat Penelitian

Manfaat yang akan diperoleh dalam penelitian ini:

Mendapatkan informasi indeks validitas yang paling baik berdasarkan nilai CTM terkecil dalam menentukan kluster optimal dan memberikan informasi stabilitas pada kluster berdasarkan koefisien *Jaccard*.



BAB II

TINJAUAN PUSTAKA

2.1 Analisis Biplot

Analisis biplot adalah representasi grafis dari informasi dalam matriks data $n \times p$. Bi mengacu pada dua jenis informasi yang terkandung dalam data (Johnson dan Wichern, 2007). Analisis biplot pertama kali di perkenalkan oleh Gabriel (1971). Biplot adalah salah satu cara menggambarkan data yang ada pada tabel ringkasan dalam grafik berdimensi dua, maka ciri-ciri variabel dan objek pengamatan serta posisi relatif antara objek pengamatan dengan variabel dapat dianalisis. Informasi dari tampilan biplot adalah (Srinadi dan Sumarjaya, 2016):

- Kedekatan antar objek, digunakan untuk melihat kemiripan karakteristik antar objek. Dua objek dengan karakteristik sama digambarkan sebagai dua titik dengan posisi berdekatan.
- Keragaman variabel, digunakan untuk melihat apakah terdapat variabel dengan keragaman yang sama untuk setiap objek. Variabel yang mempunyai keragaman yang kecil digambarkan sebagai vektor yang pendek, sedangkan variabel dengan keragaman besar digambarkan sebagai vektor panjang.
- Korelasi antar variabel, digunakan untuk mengetahui pengaruh satu variabel dengan variabel lain. Dua variabel yang memiliki korelasi positif akan digambarkan sebagai dua garis dengan arah yang sama atau membentuk sudut yang lancip. Dua variabel dengan korelasi negatif digambarkan sebagai dua garis dengan arah berlawanan atau membentuk sudut tumpul dan dua variabel yang tidak berkorelasi membentuk sudut siku-siku.
- Nilai variabel pada suatu objek, untuk melihat keunggulan dari setiap objek. Objek yang terletak searah dengan arah vektor variabel dapat dikatakan objek tersebut mempunyai nilai di atas rata-rata. Objek yang terletak berlawanan arah dengan arah vektor variabel dikatakan objek tersebut memiliki nilai di bawah rata-rata.

Perhitungan analisis biplot didasarkan pada *Singular Value Decomposition* (SVD). Matriks \mathbf{X} berukuran $n \times p$ di mana n adalah

objek dan p adalah variabel. Menurut Johnson dan Wichern (2007) penguraian nilai singular dapat dituliskan sebagai berikut:

$$\mathbf{X} = \mathbf{U} \mathbf{L} \mathbf{V}' \quad (2.1)$$

di mana

\mathbf{U} : matriks dengan kolom yang berkaitan dengan vektor eigen dari $\mathbf{X}'\mathbf{X}$ yang berukuran $n \times r$

\mathbf{L} : matriks diagonal $(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_p})$ berupa akar dari nilai eigen.

\mathbf{V} : matriks dengan kolom berupa vektor eigen dari $\mathbf{X}'\mathbf{X}$ yang berukuran $p \times r$

Menurut Jolliffe (1986) dimisalkan $\mathbf{G} = \mathbf{U}\mathbf{L}^\alpha, \mathbf{H}' = \mathbf{L}^{1-\alpha}\mathbf{V}'$ dengan $0 \leq \alpha \leq 1$ sehingga menjadi

$$\begin{aligned} \mathbf{GH}' &= \mathbf{U}\mathbf{L}^\alpha\mathbf{L}^{1-\alpha}\mathbf{V}' \\ &= \mathbf{U}\mathbf{L}\mathbf{V}' = \mathbf{X} \end{aligned} \quad (2.2)$$

Pengambilan nilai α berguna dalam interpretasi biplot yaitu $\alpha = 0$ atau $\alpha = 1$. Jika $\alpha = 0$ diperoleh $\mathbf{G} = \mathbf{U}\mathbf{L}^\alpha$ dan $\mathbf{H}' = \mathbf{L}^{1-\alpha}\mathbf{V}'$ sehingga di peroleh:

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= (\mathbf{GH}')'(\mathbf{GH}') \\ &= \mathbf{HG}'\mathbf{GH}' \\ &= \mathbf{HU}'\mathbf{UH}' \\ &= \mathbf{HH}' \end{aligned} \quad (2.3)$$

diperoleh varian dan kovarian dari \mathbf{X} adalah :

$$\mathbf{S} = \frac{1}{(n-1)} \mathbf{X}'\mathbf{X} \quad (2.4)$$

2.2 Analisis Klaster

Analisis klaster adalah teknik pengklasteran multivariat yang digunakan untuk mengklasifikasi objek atau individu ke dalam klaster yang sama berdasarkan karakteristiknya (Hair dkk., 2010). Analisis klaster dapat dikatakan sebagai analisis sederhana karena tidak memiliki asumsi. Pengklasteran dilakukan atas dasar persamaan jarak (Johnson dan Wichern, 2007). Tujuan analisis klaster adalah mengklasterkan objek-objek yang memiliki karakteristik yang sama ke dalam klaster yang sama. Hasil pengklasteran kumpulan objek akan menunjukkan homogenitas yang tinggi antar anggota dalam satu klaster (*within-cluster*) dan

heterogenitas yang tinggi kluster yang satu dengan kluster yang lain (*between-cluster*).

Secara garis besar algoritma kluster dalam mengklasifikasi terdapat dua kategori, yaitu metode hirarki dan non hirarki. Metode hirarki merupakan metode pengklasteran terstruktur dan bertahap berdasarkan kemiripan sifat antar objek. Sedangkan metode non hirarki digunakan dengan menentukan banyaknya kluster yang diinginkan.

2.2.1 Konsep Kemiripan

Konsep kemiripan adalah hal yang mendasar dalam analisis kluster. Kemiripan merupakan ukuran korespondensi antar objek di semua karakteristik yang digunakan dalam analisis. Terdapat tiga metode yang diterapkan yaitu (Hair dkk., 2010):

1. Ukuran Korelasi

Ukuran korelasi dapat diukur dengan menggunakan koefisien korelasi antar pasangan objek yang diukur dalam variabel. Ukuran korelasi digunakan apabila data bertipe metrik (data interval dan rasio). Ukuran korelasi jarang digunakan karena penekanan pada nilai suatu pola tertentu.

2. Ukuran Jarak

Ukuran jarak digunakan apabila data bertipe metrik (data interval dan rasio). Ukuran jarak merupakan ukuran kemiripan yang paling sering digunakan. Semakin tinggi nilai jarak menunjukkan kemiripan antar objek yang lebih rendah.

3. Ukuran Asosiasi

Ukuran asosiasi digunakan untuk membandingkan objek yang karakteristiknya diukur pada data bertipe nonmetrik (data nominal atau ordinal). Misalnya, responden dapat menjawab ya atau tidak pada sebuah pertanyaan. Ukuran asosiasi dapat menilai tingkat persetujuan atau kecocokan antara tiap pasangan responden.

2.2.2 Jarak *Euclidean*

Jarak *Euclidean* adalah ukuran jarak yang sering digunakan dan juga disebut sebagai jarak garis lurus. Jarak *Euclidean* mengukur jumlah kuadrat perbedaan nilai pada masing-masing variabel. Jarak *Euclidean* didefinisikan sebagai berikut (Johnson dan Wichern, 2007):

$$d_{ik} = \sqrt{\sum_{j=1}^p (x_{ij} - x_{kj})^2} \quad (2.5)$$

di mana,

d_{ik} : jarak antara objek ke- i dengan objek ke- k

p : banyaknya variabel

x_{ij} : nilai pengamatan ke- i pada variabel ke- j

x_{kj} : nilai pengamatan ke- k pada variabel ke- j

2.2.3 Metode Hirarki

Metode hirarki adalah pengklasteran terstruktur dan bertahap berdasarkan kemiripan sifat antar objek di mana banyaknya klaster belum diketahui. Metode hirarki akan membentuk semacam pohon atau dendogram dengan tingkatan antar objek, dari yang paling mirip sampai dengan yang paling tidak mirip. Secara umum terdapat dua cara pengklasteran metode hirarki, yaitu dengan cara penggabungan (*agglomerative*) dan cara pemisahan (*divisive*). Banyaknya klaster menggunakan metode penggabungan didapatkan dengan cara menggabungkan objek-objek secara bertahap sedangkan metode pemisahan langkahnya berlawanan dengan metode penggabungan. Pada metode pemisahan, semua objeknya dianggap berasal dari satu klaster besar, kemudian dilihat perbedaan antar objek. Objek yang berbeda akan dikeluarkan dari klaster awal dan seterusnya secara bertahap sehingga akan terbentuk kelompok terakhir yang beranggotakan satu objek saja.

Pembentukan klaster metode hierarki yang sering digunakan adalah algoritma penggabungan (*agglomerative*). Menurut Hair dkk., (2010), algoritma penggabungan (*agglomerative*) dibagi menjadi beberapa metode *linkage* sebagai berikut:

1. *Single Linkage*

Metode ini menggunakan jarak minimum. Jarak antar pasang objek diklasterkan ke dalam klaster yang memiliki kesamaan. Apabila objek memiliki jarak terdekat dengan objek lain, maka objek tersebut dapat digabung ke dalam satu klaster tersebut. Jika jarak antar (UV) dan klaster W didefinisikan sebagai berikut:

$$d_{(UV)W} = \min(d_{UW}, d_{VW}) \quad (2.6)$$

Nilai d_{UW} dan d_{VW} adalah jarak terdekat antara kluster U dan W dan juga kluster V dan W

2. *Complete Linkage*

Metode ini hampir sama dengan *single linkage*, yang membedakan metode *complete linkage* menggunakan jarak maksimum. Pada setiap tahap jarak antar kluster ditentukan oleh jarak antara kedua elemen dari setiap kluster yang paling jauh. Jika jarak antar (UV) dan kluster W didefinisikan sebagai berikut:

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \quad (2.7)$$

3. *Average Linkage*

Metode *average linkage* menggunakan rata-rata jarak antara semua pasangan objek sebagai jarak antara dua kluster. Metode ini dianggap lebih stabil dan tidak bias. *Average linkage* cenderung bergabung dengan kluster yang memiliki varians kecil (Everitt dkk., 2011). Jika jarak antar (UV) dan kluster W didefinisikan sebagai berikut:

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)} N_W} \quad (2.8)$$

di mana d_{ik} adalah jarak di antara objek i dalam kluster (UV) dan objek k dalam kluster W . $N_{(UV)}$ dan N_W adalah banyaknya anggota dalam kluster (UV) dan W , berturut-turut.

2.2.4 Dendogram

Dendogram merupakan gambar grafis dari proses pengklasteran. Jarak antar kombinasi lebih dekat menunjukkan homogenitas yang lebih besar. Dendogram adalah representase secara matematis dan visual dari prosedur pengklasteran yang dilakukan dengan menggunakan analisis kluster. Bentuk dendogram identik dengan diagram pohon. Menurut Hair dkk., (2010) dendogram umumnya menampilkan pola yang sama seperti yang ditunjukkan pada *agglomeration*. Dendogram akan dipotong untuk mengetahui banyaknya kluster yang terbentuk dari selisih terpanjang.

2.3 Indeks Validitas Klaster

Validitas klaster adalah suatu teknik untuk mendapatkan banyaknya klaster yang sesuai. Salah satu masalah penting dalam analisis klaster adalah evaluasi hasil pengklasteran untuk menemukan banyaknya klaster yang optimal. Indeks validitas klaster menunjukkan kebaikan hasil analisis klaster.

2.3.1 Gap Statistic

Menurut Tibshirani dkk., (2001) *gap statistic* adalah estimasi banyaknya klaster optimal. Misalkan X_{ij} adalah pengamatan pada objek ke- i dan variabel ke- j . objek data diklasterkan menjadi k klaster yaitu C_1, C_2, \dots, C_k dengan C_r merupakan pengamatan pada klaster ke- r dan n_r adalah banyaknya objek pada klaster ke- r . Jika d_{ik} merupakan jarak antara objek ke- i dan objek ke- k maka didefinisikan:

$$D_r = \sum_{(ik,jk)} d_{ik}$$

D_r merupakan jumlah jarak semua titik dalam klaster r .

$$W_k = \sum_{r=1}^k \frac{1}{2n_r} D_r \quad (2.9)$$

W_k merupakan jumlah kuadrat gabungan dalam k klaster.

Prosedur *gap statistic* sebagai berikut (Tibshirani dkk., 2001):

1. Mengklasterkan data dan memvariasikan banyaknya klaster mulai dari $k = 1, 2, \dots, K$ yang masing-masing memberikan ukuran dispersi dalam W_k .
2. Membuat sampel *bootstrap* dengan cara melakukan resampling dengan pengembalian sebanyak B kali menggunakan distribusi uniform pada daerah nilai hasil variabel pengamatan.
3. Mendapatkan nilai W_{kb}^* dari poin 2 dimana $b = 1, 2, \dots, B$ dan $k = 1, 2, \dots, K$.
4. Perhitungan klaster optimal dengan *gap statistic*:

$$Gap(k) = \left(\frac{1}{B}\right) \sum_b \log(W_{kb}^*) - \log(W_k) \quad (2.10)$$

5. Menghitung standar deviasinya yaitu:

$$sd_k = [(\frac{1}{B}) \sum_b \{\log(W_{kb}^*) - I\}^2]^{1/2} \quad (2.11)$$

$$\text{di mana } I = (\frac{1}{B}) \sum_b \log(W_{kb}^*) \quad (2.12)$$

6. Menghitung s_k dengan rumus

$$s_k = sd_k \sqrt{(1 + 1/B)} \quad (2.13)$$

7. Menghitung nilai *diffu*

$$Diffu = Gap(k) - Gap(k + 1) - s_{k+1} \quad (2.14)$$

8. Banyaknya kluster optimal diestimasi melalui nilai k terkecil sehingga $Gap(k) \geq Gap(k + 1) - s_{k+1}$ dan juga kluster optimal dilihat berdasarkan nilai $diffu \geq 0$.

2.3.2 Indeks *Silhouette*

Silhouette adalah cara lain yang dikenal untuk memperkirakan banyaknya kluster dalam satu kumpulan data. Indeks *Silhouette* diperkenalkan oleh Rousseeuw pada tahun 1987, di mana indeks tersebut dapat dihitung menggunakan persamaan (Charrod dkk., 2014):

$$S = \frac{\sum_{i=1}^n S(i)}{n}$$

dengan:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.15)$$

$a(i) : \frac{\sum dik}{n_r - 1}$, jarak rata-rata antara objek ke- i dengan seluruh pengamatan lainnya yang berada pada satu kluster yang sama.

$b(i) : \min \left\{ \frac{\sum dik}{n_s} \right\}$, jarak rata-rata antara objek ke- i dengan seluruh pengamatan lainnya yang berada pada satu kluster lainnya.

Nilai maksimum dari indeks *Silhouette* digunakan untuk menentukan banyaknya kluster optimal.

2.3.3 Indeks *Davies-Bouldin*

Indeks *Davies* dan *Bouldin* adalah fungsi dari rasio penjumlahan dalam-klaster yang menyebar ke pemisahan antar-klaster. Nilai *DB* minimum merupakan banyaknya klaster terbaik. Indeks *Davies-Bouldin* menggunakan persamaan sebagai berikut (Charrrd dkk., 2014):

$$DB = \frac{1}{K} \sum_{k=1}^K \max_{i \neq k} \left(\frac{\delta_i + \delta_k}{d_{ik}} \right) \quad (2.16)$$

di mana,

K : banyaknya klaster

$d_{ik} = \sqrt{\sum_{j=1}^p |c_{ij} - c_{kj}|^v}$ jarak antara *centroid* klaster c_i dan c_k (untuk $v = 2$ dan d_{ik} adalah jarak *Euclidean*).

$\delta_i = \sqrt{\frac{1}{ni} \sum_{i \in c_i} \sum_{j=1}^p |x_{ij} - c_{ij}|^u}$ ukuran dispersi dari klaster c_i (untuk $u = 2$ δ_i adalah standar deviasi dari jarak objek di klaster c_i ke *centroid* klaster).

2.4 Cluster Tightness Measure (CTM)

Cluster Tightness Measure (CTM) adalah ukuran kebaikan dari hasil pengklasteran berdasarkan simpangan baku setiap variabel pada masing-masing klaster, yaitu dengan rumus sebagai berikut:

$$CTM = \frac{1}{K} \sum_{k=1}^K \left(\frac{1}{p} \sum_{j=1}^p \frac{\sigma_j^k}{\sigma_j^n} \right) \quad (2.17)$$

di mana,

σ_j^k : simpangan baku pada klaster ke- k untuk variabel ke- j

σ_j^n : simpangan baku seluruh data untuk variabel ke- j

p : banyaknya variabel

K : banyaknya klaster

Hasil pengklasteran dengan nilai CTM terkecil merupakan kelompok yang dikatakan baik (Epps dan Ambikairajah, 2008).

2.5 Stabilitas Struktur Klaster

Stabilitas dapat diartikan tidak adanya perubahan banyaknya klaster jika kumpulan data diubah dengan cara yang *non-essential*.

Perubahan data asli dengan cara yang *non-essential* dilakukan dengan menggunakan metode *bootstrap*. Koefisien *Jaccard* merupakan kardinal dari irisan dua kluster dibagi dengan kardinal gabungan dua kluster tersebut. Koefisien Skema evaluasi stabilitas kluster menggunakan nilai maksimum koefisien *Jaccard* antara masing-masing kluster asli dengan kluster *bootstrap*. Rata-rata *bootstrap* dari nilai maksimum koefisien *Jaccard* merupakan derajat kestabilan untuk mengevaluasi banyak kluster yang terbentuk. Evaluasi stabilitas kluster menggunakan maksimum koefisien *Jaccard* yang didefinisikan sebagai berikut (Hennig dkk., 2016):

$$s_k^b = \max_{1 \leq k' \leq K} \frac{|C_k \cap C_{k'}^b|}{|C_k \cup C_{k'}^b|} \quad (2.18)$$

di mana,

C_k : kluster dari data asli

$C_{k'}^b$: kluster dari data *bootstrap*

Indeks stabilitas kluster didefinisikan sebagai berikut:

$$s_k = \frac{1}{B} \sum_{b=1}^B s_k^b \quad (2.19)$$

Nilai stabilitas kluster $0 \leq s_k \leq 1$. Semakin dekat s_k dengan 1 maka semakin tinggi stabilitasnya. Semakin tingginya stabilitas kluster semakin tidak mudah banyaknya kluster yang terbentuk berubah.



BAB III METODE PENELITIAN

3.1 Data dan Sumbernya

Data 1 yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) tentang jumlah kasus penyakit kabupaten/kota tahun 2016 (BPS Provinsi Jawa Timur, 2017). Ukuran data sebanyak 38 yang menunjukkan banyaknya kabupaten/kota di Jawa Timur tahun 2016. Variabel yang digunakan dalam penelitian ini sebanyak 6 variabel prediktor dengan objek yang diteliti adalah data penyakit Kabupaten/ Kota di Jawa Timur tahun 2016.

Data 2 yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Sukmawanti (2017) tentang analisis klaster dengan metode hirarki untuk pengklasteran Kabupaten/Kota di Provinsi Sulawesi Selatan berdasarkan indikator makro ekonomi. Variabel yang digunakan dalam penelitian tersebut merupakan hasil dari keseluruhan variabel dari data pertumbuhan ekonomi, kontribusi terhadap provinsi dan Indeks Pembangunan Manusia.

Data 3 yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Abdurrahman (2016) tentang analisis klaster untuk mengklasifikasikan sektor pertanian berdasarkan data jumlah rumah tangga usaha pertanian di Indonesia.

Data 4 yang digunakan dalam penelitian ini adalah data sekunder diperoleh dari Badan Pusat Statistik (BPS) tentang jumlah rasio murid dan guru kabupaten/kota tahun 2016 (BPS Provinsi Jawa Timur, 2017).

2.3 Metode yang Digunakan

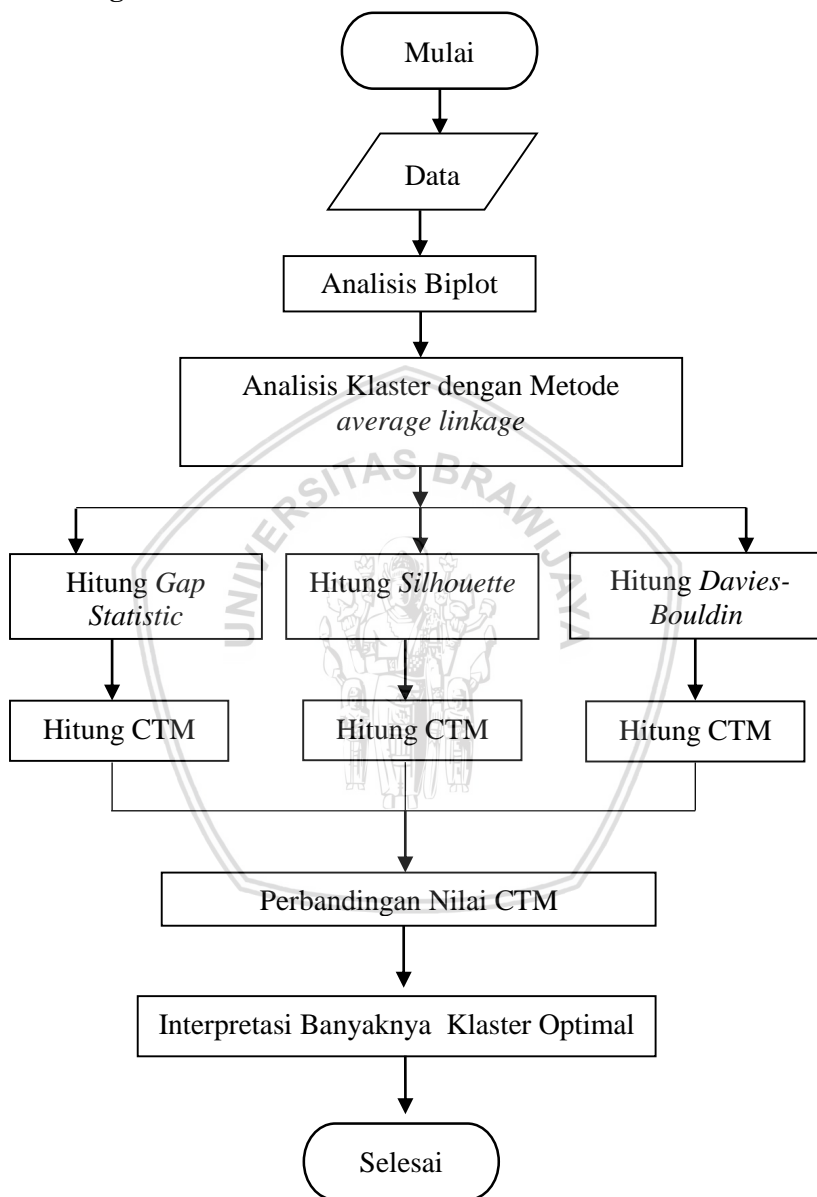
Tahapan analisis yang dilakukan untuk menyelesaikan permasalahan dalam penelitian ini adalah:

1. Melakukan analisis biplot untuk masing-masing data
2. Melakukan analisis klaster untuk menentukan banyaknya klaster (k) pada kumpulan data asli.
 - a. Menganalisis dengan jarak *Euclidean* dapat dilihat pada persamaan (2.5).
 - b. Menganalisis dengan metode *Average linkage* dapat dilihat pada persamaan (2.8).

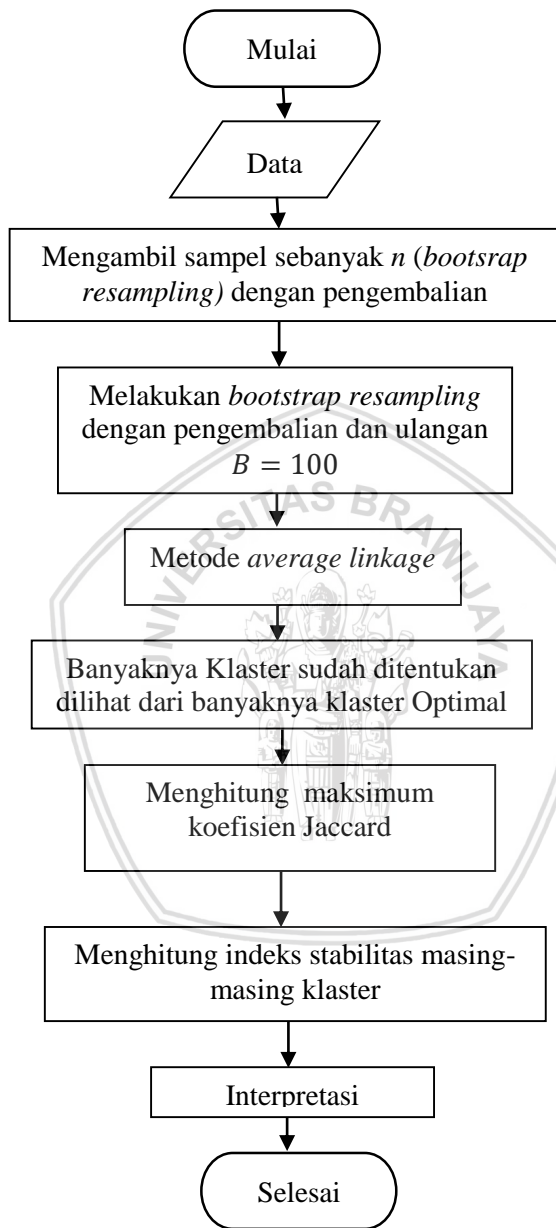
- c. Menentukan kluster optimal menggunakan *Gap Statistic*, indeks *Silhouette*, indeks *Davies-Bouldin*.
- d. Menghitung CTM masing-masing metode dapat dilihat pada persamaan (2.17).
3. Struktur Stabilitas Kluster
 - i. Mengambil sampel sebanyak n (*bootstrap resampling*).
 - ii. Melakukan analisis kluster.
 - a. Menganalisis dengan jarak *Euclidean*.
 - b. Menganalisis dengan metode *Average linkage*.
 - c. Banyaknya kluster sudah ditentukan dilihat dari banyaknya kluster optimal.
 - iii. Menghitung maksimum koefisien *Jaccard* dapat dilihat pada persamaan (2.18).
 - iv. Menghitung indeks stabilitas masing-masing kluster dapat dilihat pada persamaan (2.19).
 - v. Melakukan langkah *item* (i) sampai (iv) sebanyak B kali mendapatkan $(C_{k'}^2, C_{k'}^3, \dots, C_{k'}^B)$. Dalam penelitian ini digunakan nilai $B = 100$.

Software yang digunakan pada penelitian ini adalah Microsoft Excel dan Program R. Microsoft Excel untuk menghitung nilai CTM masing-masing indeks validitas. Program R untuk menghitung indeks validitas dan indeks stabilitas.

3.3 Diagram Alir



Gambar 3.1 Diagram Alir Banyaknya Kluster Optimal



Gambar 3.2 Diagram Alir Stabilitas Klaster

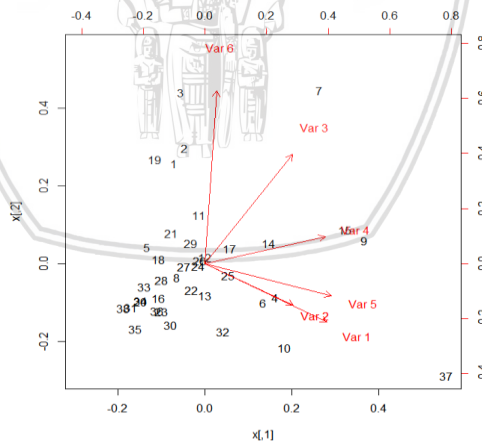
BAB IV HASIL DAN PEMBAHASAN

4.1 Hasil Analisis Biplot

Analisis biplot merupakan analisis deskriptif multivariat yang menyajikan informasi secara bersama-sama sejumlah objek pengamatan dan beberapa variabel dari suatu matriks data dalam suatu plot dimensi dua.

1. Data 1

Pada data jumlah penyakit di Provinsi Jawa Timur terdapat korelasi positif yang cukup tinggi antar variabel HIV/AIDS dan variabel IMS, korelasi ditunjukkan oleh besar sudut yang dibentuk oleh dua vektor variabel. Sudut lancip menunjukkan korelasi yang positif, sudut tumpul menunjukkan korelasi yang negatif dan sudut siku-siku menunjukkan tidak ada korelasi antara kedua variabel. Pada data ini terdapat objek yang posisinya berjauhan dengan objek lain dan dengan arah vektor variabel yang berbeda. Nilai Keragaman dapat dilihat dari panjang vektor variabel yang dibentuk, semakin panjang vektor menunjukkan tingkat keragaman yang semakin besar.

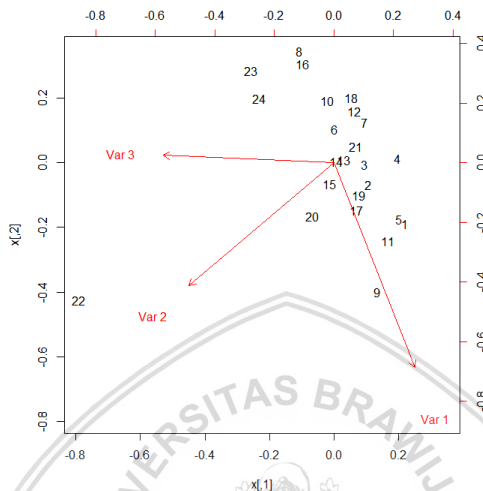


Gambar 4.1 Grafik Analisis Biplot Data 1

2. Data 2

Pada data indikator makro ekonomi di Provinsi Sulawesi Selatan tidak terdapat korelasi yang cukup tinggi antar variabel. Pada data ini

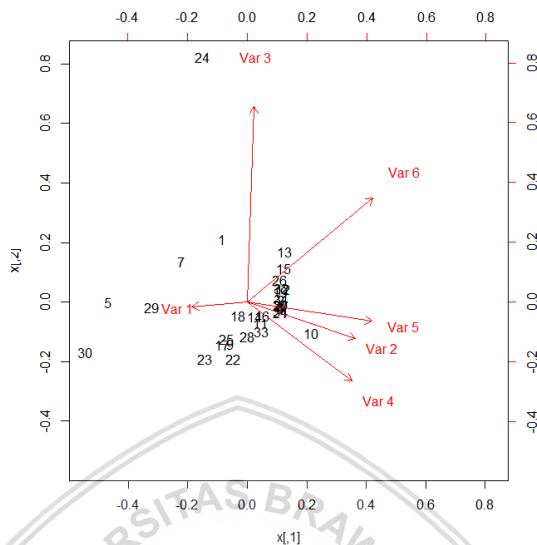
terdapat objek yang posisinya berjauhan dengan objek lain dan dengan arah vektor variabel yang sama.



Gambar 4.2 Grafik Analisis Biplot Data 2

3. Data 3

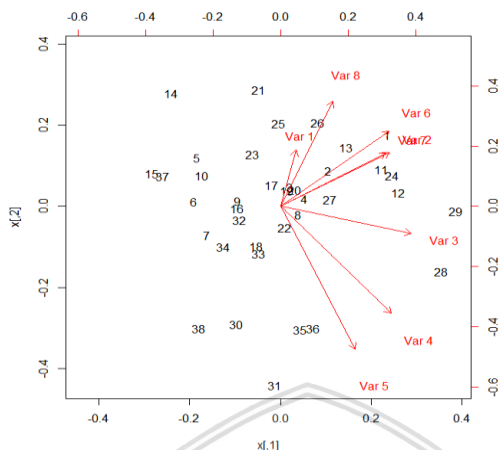
Pada data jumlah rumah tangga usaha pertanian yang melakukan pengolahan hasil pertanian menurut Provinsi di Indonesia dan subsektor tidak terdapat korelasi yang cukup tinggi antar variabel. Pada data ini tidak terdapat objek yang posisinya berjauhan dengan objek lain dan dengan arah vektor variabel yang sama.



Gambar 4.3 Grafik Analisis Biplot Data 3

4. Data 4

Pada data jumlah rasio murid dan guru di Provinsi Jawa Timur terdapat korelasi positif yang cukup tinggi antar variabel rasio guru SD dan variabel rasio guru MTS. Pada data ini terdapat objek yang posisinya berjauhan dengan objek lain dan dengan arah vektor variabel yang sama.



Gambar 4.4 Grafik Analisis Biplot Data 4

4.2 Statistika Deskriptif

Dalam statistika langkah awal sebelum melakukan analisis yang lebih kompleks adalah melakukan analisis deskriptif. Tujuan dari analisis ini adalah untuk memberikan informasi atau gambaran awal tentang variabel yang akan dianalisis.

1. Data 1

Hasil analisis deskriptif untuk dari data jumlah kasus penyakit Kabupaten/Kota di Jawa Timur tahun 2016.

Tabel 4.1 Statistika Deskriptif Data 1

Variabel	N	Mini mum	Maxi mum	Rata- Rata	Standar Deviasi	Range	Kuartil 1	Kuartil 3
HIV/AIDS	38	2	953	173.1	191.49	951	45.5	220
IMS	38	0	2200	293.5	567.35	2200	2.25	206.5
DBD	38	24	1708	666.8	411.07	1684	315.2	979.2
Diare	38	1999	66751	22770	17838	64752	8973	30697
TB	38	177	5428	1249	999.12	5251	625	1752.8
Malaria	38	0	57	9.237	14.23	57	0.25	11.75

Berdasarkan tabel 4.1 terlihat bahwa rata-rata kasus penyakit HIV/AIDS dari 38 Kabupaten/Kota di Jawa Timur sebesar 173.1 kasus. Kasus penyakit terendah sebesar 2 kasus dan penyakit

tertinggi sebesar 953 kasus dan dengan range sebesar 951. Nilai standar deviasi sebesar 191.49.

Rata-rata kasus penyakit IMS sebesar 293.53 kasus. Kasus penyakit IMS terendah adalah 0 dan penyakit tertinggi sebesar 2200 kasus dan dengan range sebesar 2200. Nilai standar deviasi sebesar 567.35.

Rata-rata kasus penyakit DBD sebesar 666.8 kasus. Kasus penyakit DBD terendah sebesar 24 kasus dan penyakit tertinggi sebesar 1708 kasus dan dengan range sebesar 1684. Nilai standar deviasi sebesar 411.07.

Rata-rata kasus penyakit Diare sebesar 22770 kasus. Kasus penyakit Diare terendah sebesar 1999 kasus dan penyakit tertinggi sebesar 66751 kasus. Nilai standar deviasi sebesar 17838.

Rata-rata kasus penyakit TB sebesar 1249.4 kasus. Kasus penyakit TB terendah sebesar 177 kasus dan penyakit tertinggi sebesar 5428 kasus. Nilai standar deviasi sebesar 999.12.

Rata-rata kasus penyakit malaria sebesar 9.237 kasus. Kasus penyakit malaria terendah sebesar 0 kasus dan penyakit tertinggi sebesar 57 kasus. Nilai standar deviasi sebesar 14.23.

2. Data 2

Hasil analisis deskriptif untuk dari data berdasarkan indikator ekonomi Kabupaten/Kota di Sulawesi Selatan tahun 2014.

Tabel 4.2 Statistika Deskriptif Data 2

Variabel	N	Mini mum	Maxi mum	Rata- Rata	Standar Deviasi	Range	Kuartil 1	Kuartil 3
Pertumbuhan Ekonomi	24	5.23	10.16	7.622	1.172	4.93	6.7	8.5
Kontribusi Terhadap Provinsi	24	1.16	33.52	4.166	6.459	32.36	1.6	3.9
Indeks Pembangunan Manusia	24	61.45	79.35	67.33	4.288	17.90	65	68.3

Berdasarkan tabel 4.2 terlihat bahwa rata-rata pertumbuhan ekonomi dari 24 Kabupaten/Kota di Sulawesi Selatan sebesar 7.622 persen. Pertumbuhan ekonomi terendah sebesar 5.23 persen dan

pertumbuhan ekonomi tertinggi sebesar 10.16 persen dan dengan range sebesar 4.93 persen. Nilai standar deviasi sebesar 1.172.

Rata-rata kontribusi terhadap provinsi sebesar 4.166 persen. Kontribusi terhadap provinsi terendah sebesar 1.16 persen dan kontribusi terhadap provinsi tertinggi sebesar 33.52 persen dan dengan range sebesar 32.26 persen. Nilai Standar deviasi sebesar 6.456.

Rata-rata indeks pembangunan manusia sebesar 67.33 persen. Indeks pembangunan manusia terendah sebesar 61.45 persen dan indeks pembangunan manusia tertinggi sebesar 79.35 persen dan dengan range sebesar 17.90 persen. Nilai standar deviasi sebesar 4.288.

3. Data 3

Hasil analisis deskriptif untuk data jumlah rumah tangga usaha pertanian di Indonesia tahun 2013.

Tabel 4.3 Statistika Deskriptif Data 3

Variabel	N	Mini mum	Maxi mum	Rata- Rata	Standar Deviasi	Range	Kuartil 1	Kuartil 3
Tanaman Pangan	34	1.12	655	46.05	117.02	653.8	3.82	24.8
Hortikular	34	1.058	802	103.4	222.86	800.9	2.34	19.6
Perkebunan	34	1.295	679	54.06	117.77	677.7	8.36	45.8
Peternakan	34	1.015	898	174.8	269.14	896.9	1.75	310.8
Perikanan	34	1.297	994	115.5	239.9	992.7	3.07	92.2
Kehutanan	34	0	882	138.5	285.48	882	2.42	29.3

Berdasarkan tabel 4.3 terlihat bahwa rata-rata subsektor tanaman pangan di Indonesia sebesar 46.05 pelaku usaha. Subsektor tanaman pangan terendah sebesar 1.12 pelaku usaha dan tanaman pangan tertinggi sebesar 655 pelaku usaha dan dengan range sebesar 653.8. Nilai standar deviasi sebesar 117.02.

Rata-rata subsektor hortikular sebesar 103.4 pelaku usaha. Subsektor hortikular terendah sebesar 1.058 pelaku usaha dan subsektor hortikular tertinggi sebesar 802 pelaku usaha dan dengan range sebesar 800.9. Nilai standar deviasi sebesar 222.86.

Rata-rata subsektor perkebunan sebesar 54.06 pelaku usaha. Subsektor perkebunan terendah sebesar 1.295 pelaku usaha dan

subsektor perkebunan tertinggi sebesar 679 pelaku usaha dan dengan range sebesar 677.7. Nilai standar deviasi sebesar 117.77.

Rata-rata subsektor peternakan sebesar 174.8 pelaku usaha. Subsektor peternakan terendah sebesar 1.015 pelaku usaha dan subsektor peternakan tertinggi sebesar 898 pelaku usaha dan dengan range sebesar 896.9. Nilai standar deviasi sebesar 269.14.

Rata-rata subsektor perikanan sebesar 115.585 pelaku usaha. Subsektor perikanan terendah sebesar 1.297 pelaku usaha dan subsektor perikanan tertinggi sebesar 994 pelaku usaha dan dengan range sebesar 992.7. Nilai standar deviasi 239.9.

Rata-rata subsektor kehutanan sebesar 138.5 pelaku usaha. Subsektor kehutanan terendah sebesar 0 pelaku usaha dan subsektor kehutanan tertinggi sebesar 882 pelaku usaha dan dengan range sebesar 882. Nilai standar deviasi sebesar 285.48.

4. Data 4

Hasil analisis deskriptif untuk dari data jumlah rasio murid dan guru di Jawa Timur tahun 2016.

Tabel 4.4 Statistika Deskriptif Data 4

Variabel	N	Mini mum	Maxi mum	Rata- Rata	Standar Deviasi	Range	Kuartil 1	Kuartil 3
TK	38	2.05	38.15	16.38	5.86	36.10	12.8	17.6
SD	38	7.76	20.27	14.70	3.36	12.51	11.9	17.2
SMP	38	6.06	18.45	12.31	2.82	12.39	10.4	14.2
SMA	38	3.84	20.79	11.40	3.73	16.95	8.9	13.7
SMK	38	3.26	15.96	10.04	2.96	12.70	8.2	12.2
MI	38	5.58	18.93	12.13	3.29	13.35	9.9	14.3
MTS	38	5.51	19.63	11.52	3.09	14.12	9.1	13.1
MA	38	1.26	14.73	9.156	2.52	13.47	7.9	10.5

Berdasarkan tabel 4.4 terlihat bahwa rata-rata rasio murid dan guru TK sebesar 16.38. Rasio murid dan guru TK terendah sebesar 2.05 dan rasio murid dan guru tertinggi sebesar 38.15 dan dengan range sebesar 36.10. Nilai standar deviasi sebesar 5.86.

Rata-rata rasio murid dan guru SD sebesar 14.70. Rasio murid dan guru SD terendah sebesar 7.76 dan rasio murid dan guru SD tertinggi sebesar 20.27 dan dengan range sebesar 12.51. Nilai standar deviasi sebesar 3.36.

Rata-rata rasio murid dan guru SMP sebesar 12.31. Rasio murid dan guru SMP terendah sebesar 6.06 dan rasio murid dan guru SMP tertinggi sebesar 18.455 dan dengan range sebesar 12.39. Nilai standar deviasi sebesar 2.82.

Rata-rata rasio murid dan guru SMA sebesar 11.40. Rasio murid dan guru SMA terendah sebesar 3.84 dan rasio murid dan guru SMA tertinggi sebesar 20.79 dan dengan range sebesar 16.95. Nilai standar deviasi sebesar 3.73.

Rata-rata rasio murid dan guru SMK sebesar 10.04. Rasio murid dan guru SMK terendah sebesar 3.26 dan rasio murid dan guru SMK tertinggi sebesar 15.96 dan dengan range sebesar 12.70. Nilai standar deviasi sebesar 2.96.

Rata-rata rasio murid dan guru MI sebesar 12.13. Rasio murid dan guru MI terendah sebesar 5.58 dan rasio murid dan guru MI tertinggi sebesar 18.93 dan dengan range sebesar 13.35. Nilai standar deviasi sebesar 3.29.

Rata-rata rasio murid dan guru MTS sebesar 11.52. Rasio murid dan guru MTS terendah sebesar 5.51 dan rasio murid dan guru MTS tertinggi sebesar 19.63 dan dengan range sebesar 14.12. Nilai standar deviasi sebesar 3.09.

Rata-rata rasio murid dan guru MA sebesar 9.156. Rasio murid dan guru MA terendah sebesar 1.26 dan rasio murid dan guru MA tertinggi sebesar 14.73 dan dengan range sebesar 13.47. Nilai standar deviasi sebesar 2.52.

4.3 Hasil Analisis Klaster

Metode *linkage* yang digunakan adalah *Average Linkage* dengan menggunakan ukuran jarak *Euclidean* pada persamaan (2.5). Penentuan banyaknya klaster optimal berdasarkan 3 metode indeks validitas yaitu *Gap Statistic*, *Silhouette*, dan *Davies-Bouldien*.

4.4 Hasil Analisis Indeks Validitas (IV)

4.4.1 IV untuk *Gap Statistic*

1. Data 1

Gap statistic dalam menentukan banyaknya klaster optimal dilihat berdasarkan dari nilai *diffu*. Perhitungan nilai *gap statistic* menggunakan persamaan (2.10) dan nilai *diffu* menggunakan persamaan (2.14).

Tabel 4.5 Nilai Indeks Validitas *Gap Statistic* Pada Data 1

<i>K</i>	<i>Gap (K)</i>	Nilai <i>Diffu</i>
2	-0.0003364	-0.2136372
3	0.4211168	-0.0497566
4	0.6547877	0.4836613
5	0.360272	-0.0040209

Tabel 4.5 merupakan hasil perhitungan nilai validitas kluster dari indeks *gap statistic*. Pada data kasus penyakit mengindikasikan bahwa banyaknya kluster optimal yang didapatkan adalah 4 kluster karena pada kluster ke-4 diperoleh nilai *diffu* ($0.4836613 \geq 0$). Dari grafik pada Lampiran 7, saat terbentuk 2 dan 3 kluster nilai *diffu* masih di bawah 0. Tetapi pada saat terbentuk 4 kluster nilai *diffu* lebih besar dari 0.

2. Data 2

Gap statistic dalam menentukan banyaknya kluster optimal dilihat berdasarkan dari nilai *diffu*. Perhitungan nilai *gap statistic* menggunakan persamaan (2.10) dan nilai *diffu* menggunakan persamaan (2.14).

Tabel 4.6 Nilai Indeks Validitas *Gap Statistic* Pada Data 2

<i>K</i>	<i>Gap (K)</i>	Nilai <i>Diffu</i>
2	1.257808	-0.0150261
3	1.605923	0.5568546
4	1.695922	0.5278796
5	1.910536	0.7059517

Tabel 4.6 merupakan hasil perhitungan nilai validitas kluster dari indeks *gap statistic*. Pada data makro ekonomi mengindikasikan bahwa banyaknya kluster optimal yang didapatkan adalah 3 kluster karena pada kluster ke-3 diperoleh nilai *diffu* ($0.5568546 \geq 0$). Dari grafik pada Lampiran 8, saat terbentuk 2 kluster nilai *diffu* masih di bawah 0. Tetapi pada saat terbentuk 3 kluster nilai *diffu* lebih besar dari 0.

3. Data 3

Gap statistic dalam menentukan banyaknya kluster optimal dilihat berdasarkan dari nilai *diffu*. Perhitungan nilai *gap statistic*

menggunakan persamaan (2.10) dan nilai *diffu* menggunakan persamaan (2.14).

Tabel 4.7 Nilai Indeks Validitas *Gap Statistic* Pada Data 3

<i>K</i>	<i>Gap (K)</i>	Nilai <i>Diffu</i>
2	0.396283	0.1638922
3	0.3301813	-0.1327707
4	0.5667653	0.9608328
5	0.5490486	0.0871941

Tabel 4.7 merupakan hasil perhitungan nilai validitas kluster dari indeks *gap statistic*. Pada data sektor pertanian mengindikasikan bahwa banyaknya kluster optimal yang didapatkan adalah 2 kluster karena pada kluster ke-2 diperoleh nilai *diffu* $(0.1638922) \geq 0$. Dari grafik pada Lampiran 9, saat terbentuk 2 kluster nilai *diffu* lebih besar dari 0.

4. Data 4

Gap statistic dalam menentukan banyaknya kluster optimal dilihat berdasarkan dari nilai *diffu*. Perhitungan nilai *gap statistic* menggunakan persamaan (2.10) dan nilai *diffu* menggunakan persamaan (2.14).

Tabel 4.8 Nilai Indeks Validitas *Gap Statistic* Pada Data 4

<i>K</i>	<i>Gap (K)</i>	Nilai <i>Diffu</i>
2	0.5718313	0.1115861
3	0.5281812	-0.1538472
4	0.7469773	0.9399786
5	0.7281538	-0.07940273

Tabel 4.8 merupakan hasil perhitungan nilai validitas kluster dari indeks *gap statistic*. Pada data rasio antara murid dan guru mengindikasikan bahwa banyaknya kluster optimal yang didapatkan adalah 2 kluster karena pada kluster ke-2 diperoleh nilai *diffu* $(0.1115861) \geq 0$. Dari grafik pada Lampiran 10, saat terbentuk 2 kluster nilai *diffu* lebih besar dari 0.

4.4.2 IV untuk *Silhouette*

1. Data 1

Berdasarkan hasil pengelompokan dengan metode *average linkage* dan menghitung nilai indeks validitas *Silhouette* menggunakan persamaan (2.15) dengan $K = 2,3,4,5$ didapatkan hasil seperti pada Tabel 4.9.

Tabel 4. 9 Nilai Indeks Validitas *Silhouette* Pada Data 1

K	<i>Silhouette</i>
2	0.7
3	0.62
4	0.61
5	0.57

Berdasarkan hasil perhitungan nilai validitas kluster dari indeks validitas *Silhouette* banyaknya kluster optimal ditunjukkan oleh nilai S yang paling mendekati satu atau nilai terbesar. Pada Tabel 4.9 didapatkan banyaknya kluster optimal yang diperoleh adalah 2 karena nilai S yang terbesar adalah 0.7.

2. Data 2

Menghitung nilai indeks validitas *Silhouette* menggunakan persamaan (2.15). Nilai indeks validitas *Silhouette* dengan $K = 2,3,4,5$ didapatkan hasil seperti pada Tabel 4.10

Tabel 4.10 Nilai Indeks Validitas *Silhouette* Pada Data 2

K	<i>Silhouette</i>
2	0.82
3	0.6
4	0.39
5	0.33

Berdasarkan hasil perhitungan nilai validitas kluster dari indeks validitas *Silhouette* banyaknya kluster optimal ditunjukkan oleh nilai S yang paling mendekati satu atau nilai terbesar. Pada Tabel 4.10 didapatkan banyaknya kluster optimal yang diperoleh adalah 2 karena nilai S yang terbesar adalah 0.82.

3. Data 3

Menghitung nilai indeks validitas *Silhouette* menggunakan persamaan (2.15). Nilai indeks validitas *Silhouette* dengan $K = 2,3,4,5$ didapatkan hasil seperti pada Tabel 4.11.

Tabel 4.11 Nilai Indeks Validitas *Silhouette* Pada Data 3

K	<i>Silhouette</i>
2	0.53
3	0.49
4	0.5
5	0.49

Berdasarkan hasil perhitungan nilai validitas kluster dari indeks validitas *Silhouette* banyaknya kluster optimal ditunjukkan oleh nilai S yang paling mendekati satu atau nilai terbesar. Pada Tabel 4.11 didapatkan banyaknya kluster optimal yang diperoleh adalah 2 karena nilai S yang terbesar adalah 0.53.

4. Data 4

Menghitung nilai indeks validitas *Silhouette* menggunakan persamaan (2.15). Nilai indeks validitas *Silhouette* dengan $K = 2,3,4,5$ didapatkan hasil seperti pada Tabel 4.12.

Tabel 4.12 Nilai Indeks Validitas *Silhouette* Pada Data 4

K	<i>Silhouette</i>
2	0.36
3	0.25
4	0.26
5	0.22

Berdasarkan hasil perhitungan nilai validitas kluster dari indeks validitas *Silhouette* banyaknya kluster optimal ditunjukkan oleh nilai S yang paling mendekati satu atau nilai terbesar. Pada Tabel 4.12 didapatkan banyaknya kluster optimal yang diperoleh adalah 2 karena nilai S yang terbesar adalah 0.36.

4.4.3 IV untuk *Davies-Bouldien*

1. Data 1

Hasil pengklasteran menggunakan *average* linkage, kemudian dihitung indeks validitas *Davies-Bouldin* dengan menggunakan persamaan (2.16).

Tabel 4.13 Nilai Indeks Validitas *Davies-Bouldin* Pada Data 1

<i>K</i>	<i>Davies-Bouldin</i>
2	0.339137
3	0.493939
4	0.478858
5	0.461908

Berdasarkan Tabel 4.13 hasil perhitungan nilai validitas kluster dari indeks validitas *Davies-Bouldin* banyaknya kluster optimal ditunjukkan oleh nilai *Davies-Bouldin* terkecil. Diperoleh banyaknya kluster optimal adalah 2 kluster dengan *Davies-Bouldin* sebesar 0.339137. Dari grafik Lampiran 11 saat terbentuk 2 kluster nilai *Davies Bouldin* lebih kecil dibandingkan dengan nilai kluster 3 sampai 5.

2. Data 2

Hasil pengklasteran menggunakan *average* linkage, kemudian dihitung indeks validitas *Davies-Bouldin*.

Tabel 4.14 Nilai Indeks Validitas *Davies-Bouldin* Pada Data 2

<i>K</i>	<i>Davies-Bouldin</i>
2	0.120282
3	0.249530
4	0.344948
5	0.563406

Berdasarkan Tabel 4.14 hasil perhitungan nilai validitas kluster dari indeks validitas *Davies-Bouldin* banyaknya kluster optimal ditunjukkan oleh nilai *Davies-Bouldin* terkecil. Diperoleh banyaknya kluster optimal adalah 2 kluster dengan *Davies-Bouldin* sebesar 0.120282. Dari grafik Lampiran 12 saat terbentuk 2 kluster nilai *Davies Bouldin* lebih kecil dibandingkan dengan nilai kluster 3 sampai 5.

3. Data 3

Hasil pengklasteran menggunakan *average* linkage, kemudian dihitung indeks validitas *Davies-Bouldin*.

Tabel 4.15 Nilai Indeks Validitas *Davies-Bouldin*

<i>K</i>	<i>Davies-Bouldin</i>
2	0.937089
3	0.624338
4	0.799833
5	0.571641

Berdasarkan Tabel 4.15 hasil perhitungan nilai validitas kluster dari indeks validitas *Davies-Bouldin* banyaknya kluster optimal ditunjukkan oleh nilai *Davies-Bouldin* terkecil. Diperoleh banyaknya kluster optimal adalah 5 kluster dengan *Davies-Bouldin* sebesar 0.571641. Dari grafik Lampiran 13 saat terbentuk 5 kluster nilai *Davies Bouldin* lebih kecil dibandingkan dengan nilai kluster 2 sampai 4.

4. Data 4

Hasil pengklasteran menggunakan *average* linkage, kemudian dihitung indeks validitas *Davies-Bouldin*.

Tabel 4.16 Nilai Indeks Validitas *Davies-Bouldin* Pada Data 4

<i>K</i>	<i>Davies-Bouldin</i>
2	0.9433812
3	0.833619
4	0.8835498
5	0.7548539

Berdasarkan Tabel 4.16 hasil perhitungan nilai validitas kluster dari indeks validitas *Davies-Bouldin* banyaknya kluster optimal ditunjukkan oleh nilai *Davies-Bouldin* terkecil. Diperoleh banyaknya kluster optimal adalah 5 kluster dengan *Davies-Bouldin* sebesar 0.7548539. Dari grafik Lampiran 14 saat terbentuk 5 kluster nilai *Davies Bouldin* lebih kecil dibandingkan dengan nilai kluster 2 sampai 4.

4.5 Hasil Perhitungan *Cluster Tightness Measure* (CTM)

Cluster Tightness Measure (CTM) merupakan untuk mengetahui metode pengklasteran terbaik, di mana semakin kecil nilai CTM maka pengklasteran semakin baik.

4.5.1 CTM Untuk *Gap statistic*

1. Data 1

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *Gap Statistic* diperoleh 4 klaster.

Tabel 4.17 Simpangan Baku *Gap Statistic* Pada Data 1

Variabel	Simpangan Baku				Simpangan Baku Data
	Klaster 1	Klaster 2	Klaster 3	Klaster 4	
X1	80.6023	141.312	298.953	120.855	191.4995
X2	204.976	723.196	63.9221	770.832	567.3556
X3	343.331	369.708	50.8904	290.552	411.0776
X4	3909.28	3312.32	206.962	2859.14	17838.13
X5	531.046	417.045	10.5639	364.116	999.1294
X6	12.0409	17.3349	0.85492	3.01188	14.2344

Berdasarkan simpangan baku pada Tabel 4.17 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan menggunakan persamaan (2.17) di mana banyaknya klaster $K = 4$ dan banyaknya variabel $p = 6$. Nilai CTM pada indeks validitas *Gap Statistic* adalah 0.55248.

2. Data 2

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *Gap Statistic* diperoleh 3 klaster.

Tabel 4.18 Simpangan Baku *Gap Statistic* Data 2

Variabel	Simpangan Baku			Simpangan Baku Data
	Klaster 1	Klaster 2	Klaster 3	
X1	1.00808	0	0.40305	1.172506
X2	0.26054	0	0.07778	6.459947
X3	0.5238	0	0.00707	4.288546

Berdasarkan simpangan baku pada Tabel 4.18 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan menggunakan persamaan (2.17) di mana banyaknya kluster $K = 3$ dan banyaknya variabel $p = 3$. Nilai CTM pada indeks validitas *Gap Statistic* adalah 0.31831.

3. Data 3

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas kluster dari *Gap Statistic* diperoleh 2 kluster.

Tabel 4.19 Simpangan Baku *Gap Statistic* Pada Data 3

Variabel	Simpangan Baku		Simpangan Baku Data
	Klaster 1	Klaster 2	
X1	121.8832	1.195677	117.0179
X2	162.887	425.8154	222.8566
X3	122.7293	3.234992	117.7716
X4	241.9738	194.7032	269.1406
X5	154.2374	413.363	239.9012
X6	236.1876	152.3297	285.4839

Berdasarkan simpangan baku pada Tabel 4.19 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan menggunakan persamaan (2.17) di mana banyaknya kluster $K = 2$ dan banyaknya variabel $p = 6$. Nilai CTM pada indeks validitas *Gap Statistic* adalah 0.31831.

4. Data 4

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *Gap Statistic* diperoleh 2 klaster.

Tabel 4.20 Simpangan Baku *Gap Statistic* Pada Data 4

Variabel	Simpangan Baku		Simpangan Baku Data
	Klaster 1	Klaster 2	
X1	3.752743	5.994172	5.859113
X2	3.298538	1.465026	3.360062
X3	2.896443	1.977347	2.823384
X4	3.753938	4.146183	3.738601
X5	3.084427	0.772938	2.966881
X6	3.390927	2.197961	3.291592
X7	3.128166	2.478124	3.088525
X8	2.521156	1.133711	2.521461

Berdasarkan simpangan baku pada Tabel 4.20 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan menggunakan persamaan (2.17) di mana banyaknya klaster $K = 2$ dan banyaknya variabel $p = 8$. Nilai CTM pada indeks validitas *Gap Statistic* adalah 1.164504.

4.5.2 CTM Untuk *Silhouette*

1. Data 1

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *silhouette* diperoleh 2 klaster.

Tabel 4.21 Simpangan Baku *Silhouette* Pada Data 1

Variabel	Simpangan Baku		Simpangan baku data
	Klaster 1	Klaster 2	
X1	130.2355	298.9535	191.4995
X2	529.4307	668.1319	567.3556
X3	358.5589	364.3202	411.0776
X4	11936.49	4276.667	17838.13
X5	506.6255	1364.665	999.1294
X6	12.4648	22.92815	14.2344

Berdasarkan simpangan baku pada Tabel 4.21 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan menggunakan persamaan (2.17) diperoleh nilai CTM pada indeks validasi *Silhouette* adalah 0.94822.

2. Data 2

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *silhouette* diperoleh 2 klaster.

Tabel 4.22 Simpanan Baku *Silhouette* Pada Data 2

Variabel	Simpangan Baku		Simpangan baku data
	Klaster 1	Klaster 2	
X1	1.197796	0	1.172506
X2	1.660779	0	6.459947
X3	3.518187	0	4.288546

Berdasarkan simpangan baku pada Tabel 4.22 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan menggunakan persamaan (2.17) diperoleh nilai CTM pada indeks validasi *Silhouette* adalah 0.349838.

3. Data 3

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *Silhouette* diperoleh 2 klaster.

Tabel 4.23 Simpanan Baku *Silhouette* Pada Data 3

Variabel	Simpangan Baku		Simpangan Baku Data
	Klaster 1	Klaster 2	
X1	121.8832	1.195677	117.0179
X2	162.887	425.8154	222.8566
X3	122.7293	3.234992	117.7716
X4	241.9738	194.7032	269.1406
X5	154.2374	413.363	239.9012
X6	236.1876	152.3297	285.4839

Berdasarkan simpangan baku pada Tabel 4.23 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan

menggunakan persamaan (2.17) diperoleh nilai CTM pada indeks validitas *Silhouette* adalah 1.253401.

4. Data 4

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *Silhouette* diperoleh 2 klaster.

Tabel 4.24 Simpangan Baku *Silhouette* Pada Data 4

Variabel	Simpangan Baku		Simpangan Baku Data
	Klaster 1	Klaster 2	
X1	3.752743	5.994172	5.859113
X2	3.298538	1.465026	3.360062
X3	2.896443	1.977347	2.823384
X4	3.753938	4.146183	3.738601
X5	3.084427	0.772938	2.966881
X6	3.390927	2.197961	3.291592
X7	3.128166	2.478124	3.088525
X8	2.521156	1.133711	2.521461

Berdasarkan simpangan baku pada Tabel 4.24 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan menggunakan persamaan (2.17) diperoleh nilai CTM pada indeks validitas *Silhouette* adalah 1.164504.

4.5.3 CTM Untuk *Davies-Bouldin*

1. Data 1

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *Davies-Bouldin* diperoleh 2 klaster.

Tabel 4.25 Simpanan Baku *Davies-Bouldin* Data 1

Variabel	Simpangan Baku		Simpangan baku data
	Klaster 1	Klaster 2	
X1	130.2355	298.9535	191.4995
X2	529.4307	668.1319	567.3556
X3	358.5589	364.3202	411.0776
X4	11936.49	4276.667	17838.13
X5	506.6255	1364.665	999.1294
X6	12.4648	22.92815	14.2344

Berdasarkan simpangan baku pada Tabel 4.25 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan menggunakan persamaan (2.17) diperoleh nilai CTM pada indeks validasi *Davies-Bouldin* adalah 0.94822.

2. Data 2

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *Davies-Bouldin* diperoleh 2 klaster.

Tabel 4.26 Simpanan Baku *Davies-Bouldin* Data 2

Variabel	Simpangan Baku		Simpangan baku data
	Klaster 1	Klaster 2	
X1	1.197796	0	1.172506
X2	1.660779	0	6.459947
X3	3.518187	0	4.288546

Berdasarkan simpangan baku pada Tabel 4.26 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan menggunakan persamaan (2.17) diperoleh nilai CTM pada indeks validasi *Davies-Bouldin* adalah 0.349838.

3. Data 3

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *Davies-Bouldin* diperoleh 2 klaster.

Tabel 4.27 Simpanan Baku *Davies-Bouldin* Data 3

Variabel	Simpangan Baku					Simpangan baku data
	Klaster 1	Klaster 2	Klaster 3	Klaster 4	Klaster 5	
X1	13.981	127.62	0	0	0.63781	117.0179
X2	1.9403	167.88	0	0	105.358	222.8566
X3	15.653	44.305	0	0	0.20576	117.7716
X4	1.6510	252.29	0	0	271.529	269.1406
X5	410.72	126.67	0	0	366.281	239.9012
X6	62.932	21.445	0	0	87.6812	285.4839

Berdasarkan simpangan baku pada Tabel 4.27 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan

menggunakan persamaan (2.17) diperoleh nilai CTM pada indeks validasi *Davies-Bouldin* adalah 1.253904.

4. Data 4

Berdasarkan proses pengklasteran menggunakan *average linkage* dan dihitung validitas klaster dari *Davies-Bouldin* diperoleh 2 klaster.

Tabel 4.28 Simpanan Baku *Davies-Bouldin* Data 4

Variabel	Simpanan Baku					Simpanan baku data
	Klaster 1	Klaster 2	Klaster 3	Klaster 4	Klaster 5	
X1	2.6504	3.0324	0	0.48790	0	5.859113
X2	1.5449	2.9347	0	1.1808	0	3.360062
X3	1.8317	2.4538	0	1.86676	0	2.823384
X4	2.8623	3.5671	0	1.94454	0	3.738601
X5	2.0580	2.8991	0	0.20506	0	2.966881
X6	0.8482	2.7601	0	0.55861	0	3.291592
X7	1.1700	2.7251	0	2.14960	0	3.088525
X8	1.2005	2.7247	0	1.54149	0	2.521461

Berdasarkan simpanan baku pada Tabel 4.28 dapat digunakan untuk menghitung *Cluster Tightness Measure* (CTM) dengan menggunakan persamaan (2.13) diperoleh nilai CTM pada indeks validasi *Davies-Bouldin* adalah 1.372708.

4.5.4 Perbandingan Hasil CTM

Hasil perbandingan indeks validasi kluster dengan *Gap Statistic*, *Silhouette* dan *Davies-Bouldin* disajikan pada Tabel 4.30:

Tabel 4.29 Nilai CTM Indeks Validitas *Gap Statistic*, *Silhouette* dan *Davies-Bouldin*

Metode	Nilai CTM Data 1	Nilai CTM Data 2	Nilai CTM Data 3	Nilai CTM Data 4
<i>Gap Statistic</i>	0.55248	0.31831	1.253401	1.164504
<i>Silhouette</i>	0.94822	0.349838	1.253401	1.164504
<i>Davies- Bouldin</i>	0.94822	0.349838	1.253904.	1.372708

Berdasarkan Tabel 4.29 nilai CTM data 1 dapat disimpulkan indeks validitas *Gap Statistic* merupakan indeks validitas yang lebih baik dalam penentuan banyaknya kluster optimal karena nilai CTM indeks validitas *Gap Statistic* lebih kecil dibandingkan dengan *Silhouette* dan *Davies-Bouldin* pada analisis kluster metode *average linkage*. Pada data 2 dapat disimpulkan indeks validitas *Gap Statistic* juga merupakan indeks validitas yang lebih baik dalam penentuan banyaknya kluster optimal karena nilai CTM indeks validitas *Gap Statistic* lebih kecil dibandingkan dengan *Silhouette* dan *Davies-Bouldin* pada analisis kluster metode *average linkage*. Pada data 3 dan 4 dapat disimpulkan indeks validitas *Gap Statistic* dan *Silhouette* merupakan indeks validitas yang lebih baik dalam penentuan banyaknya kluster optimal karena nilai CTM indeks validitas *Gap Statistic* dan *Silhouette* lebih kecil dibandingkan *Davies-Bouldin* pada analisis kluster metode *average linkage*.

4.6 Karakteristik Data

Karakteristik setiap data digunakan dilihat berdasarkan rata-rata, standar deviasi dan range. Pada data 1 berdasarkan rata-rata, nilai standar deviasi dan range bervariasi besar untuk setiap variabel. Berdasarkan grafik analisis biplot pada data 1 di antara variabel-variabel terdapat korelasi positif yang cukup tinggi dan terdapat objek yang posisinya berjauhan dengan objek lain dengan arah

vektor variabel yang berbeda. Indeks validitas terbaik untuk data 1 adalah *gap statistic*.

Pada data 2 berdasarkan rata-rata dan range bervariasi cukup besar, dan standar deviasi bervariasi kecil untuk setiap variabel. Setelah di uji indeks validitas yang terbaik adalah *gap statistic*. Berdasarkan grafik analisis biplot pada data 2 di antara variabel-variabel tidak terdapat korelasi positif yang cukup tinggi dan terdapat objek yang posisinya berjauhan dengan objek lain dengan arah vektor variabel yang sama. Indeks validitas terbaik untuk data 1 adalah *gap statistic*.

Pada data 3 berdasarkan rata-rata range bervariasi cukup besar, nilai standar deviasi bervariasi kecil untuk setiap variabel. Berdasarkan grafik analisis biplot pada data 3 di antara variabel-variabel tidak terdapat korelasi positif yang cukup tinggi dan tidak terdapat objek yang posisinya berjauhan dengan objek lain dengan arah vektor variabel yang sama. Setelah diuji indeks validitas yang terbaik adalah *gap statistic* dan *silhouette*.

Pada data 4 berdasarkan rata-rata, nilai standar deviasi dan range bervariasi kecil untuk setiap variabel. Berdasarkan grafik analisis biplot pada data 4 di antara variabel-variabel terdapat korelasi positif yang cukup tinggi dan terdapat objek yang posisinya berjauhan dengan objek lain dengan arah vektor variabel yang sama. Setelah diuji indeks validitas yang terbaik adalah *gap statistic* dan *silhouette*.

Berdasarkan keempat data yang digunakan dapat disimpulkan bahwa indeks validitas *gap statistic* memiliki karakteristik data pada kondisi data bervariasi yang besar atau kecil, pada data yang di antara variabel-variabel terdapat korelasi yang cukup tinggi atau tidak ada korelasi dan terdapat objek yang posisinya berjauhan dengan objek lain dengan arah vektor yang sama atau berberda.

Indeks validitas *silhouette* memiliki karakteristik pada kondisi data bervariasi kecil, pada data yang di antara variabel-variabel terdapat korelasi yang cukup tinggi atau tidak ada korelasi dan terdapat objek yang posisinya berjauhan dengan objek lain dengan arah vektor variabel yang sama.

Indeks validitas *Davies-Bouldin* memiliki karakteristik pada kondisi data bervariasi cukup besar, pada data yang di antara variabel-variabel tidak terdapat korelasi yang cukup tinggi dan tidak

terdapat objek yang posisinya berjauhan dengan objek lain dengan arah vektor variabel yang sama.

4.7 Hasil Analisis Struktur Stabilitas Klaster

1. Data 1

Definisi koefisien *Jaccard* nilai stabilitas klaster antara 0 sampai 1. Semakin dekat dengan 1 adalah semakin stabil klaster tersebut. Perhitungan stabilitas klaster dengan menggunakan persamaan (2.19). Berdasarkan pengamatan diperoleh nilai indeks stabilitas disajikan dalam Tabel 4.30.

Tabel 4.30 Struktur Stabilitas Klaster Data 1

Klaster	Indeks Stabilitas
1	0.88196
2	0.73024
3	0.961667
4	0.89142

Nilai indeks stabilitas untuk klaster C_3 memiliki indeks stabilitas yang sangat stabil, klaster C_1, C_4 memiliki indeks stabilitas yang stabil dan klaster C_2 memiliki indeks stabilitas yang cukup stabil.

Anggota klaster pertama : Pacitan, Blitar, Mojokerto, Nganjuk, Madiun, Magetan, Ngawi, Lamongan, Sumenep, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Batu.

Anggota klaster kedua : Ponorogo, Trenggalek, Tulungagung, Kediri, Lumajang, Situbondo, Tuban, Pamekasan.

Anggota klaster ketiga : Malang, Jember, Pasuruan, Sidoarjo, Kota Surabaya.

Anggota klaster keempat : Bangkalan, Bondowoso, Probolinggo, Jombang, Bojonegoro, Gresik, Sampang, Banyuwangi.

2. Data 2

Berdasarkan pengamatan diperoleh nilai indeks stabilitas disajikan dalam Tabel 4.31.

Tabel 4.31 Struktur Stabilitas Klaster Data 2

Klaster	Indeks Stabilitas
1	0.9251894
2	0.61
3	0.85

Nilai indeks stabilitas untuk klaster C_1 , menunjukkan bahwa klaster tersebut memiliki indeks stabilitas yang sangat tinggi, C_3 memiliki indeks stabilitas yang stabil dan klaster C_2 memiliki indeks stabilitas yang kurang stabil.

Anggota klaster pertama : Kep Selayar, Bulukumba, Banteng, Jeneponto, Takalar, Gowa, Sinjai, Maros, Pangkep, Barru, Bone, Soppeng, Wajo, Sindrap, Pinrang, Enrekang, Luwu, Tana Toraja, Luwu Utara, Luru Timur, Toraja Utara.

Anggota klaster kedua : Makassar.

Anggota klaster ketiga : Pare-Pare, Palopo.

3. Data 3

Berdasarkan pengamatan diperoleh nilai indeks stabilitas disajikan dalam Tabel 4.32.

Tabel 4.32 Struktur Stabilitas Klaster Data 3

Klaster	Indeks Stabilitas
1	0.9634048
2	0.9912377

Nilai indeks stabilitas untuk klaster C_1 dan C_2 menunjukkan bahwa klaster tersebut memiliki indeks stabilitas yang sangat tinggi.

Anggota klaster pertama : Aceh, Sumatera Utara, Sumatera Barat, Riau, Sumatera Selatan,

Bengkulu, Lampung, Kep Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi tengah, Sulawesi Selatan, Sulawesi Tenggara, Maluku, Maluku Utara, Papua Barat, Papua.

Anggota klaster kedua :Jambi, Gorontalo, Sulawesi Barat.

4. Data 4

Berdasarkan pengamatan diperoleh nilai indeks stabilitas disajikan dalam Tabel 4.33.

Tabel 4.33 Struktur Stabilitas Klaster Data 4

Klaster	Indeks Stabilitas
1	0.9571667
2	0.9901830

Nilai indeks stabilitas untuk klaster C_1 , dan C_2 menunjukkan bahwa klaster tersebut memiliki indeks stabilitas yang sangat tinggi.

Anggota klaster pertama :Pacitan, Ponogoro, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Situbondo, Probolinggo, Pasuruan, Sidoarjo, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik, Bangkalan, Sampang, Pamekasan, Sumenep, Kota Kediri, Kota Blitar, Kota Malang, Kota Madiun, Kota Surabaya, Kota Batu.

Anggota klaster kedua :Kota Probolinggo, Kota Mojokerto, Kota Pasuruan.

4.8 Interpretasi dan Memprofil Klaster

Menginterpretasikan dan memprofil klaster merupakan penjelasan karakteristik masing-masing klaster untuk menjelaskan bagaimana perbedaan pada dimensi yang relevan dan mengetahui karakteristik yang dominan pada masing-masing klaster. Nilai rata-rata objek yang terdapat dalam klaster pada setiap variabel digunakan untuk memprofil klaster. Nilai rata-rata objek dalam klaster ini dapat digunakan untuk memberikan nama atau label dibandingkan dengan nilai kuartil 1 dan kuartil 3 masing-masing variabel untuk keseluruhan data. Didefinisikan nilai rata-rata objek dalam klaster yang berada di bawah nilai kuartil 1 dapat didefinisikan rendah, nilai rata-rata yang berada di antara nilai kuartil 1 dan kuartil 3 dapat didefinisikan sedang dan nilai rata-rata objek dalam klaster yang berada di atas nilai kuartil 3 dapat didefinisikan tinggi.

1. Data 1

Tabel 4.34 Nilai Rata-Rata Objek Dalam Klaster Pada Setiap Variabel Data 1

Variabel	Klaster 1	Klaster 2	Klaster 3	Klaster 4
HIV/AIDS	98.11765	155.875	504.2	142.875
IMS	114.2353	395	741.4	293.125
DBD	447.5294	728.375	1195.2	741
Diare	83277.353	20464.88	60166.6	32391.5
TB	770.4118	1016.375	3171.6	1299
Malaria	8.117647	13.25	17.8	2.25

Berdasarkan tabel 4.34 dapat dijabarkan bahwa pada klaster 1 terdiri dari Kabupaten/Kota di Jawa Timur dengan rata-rata kasus penyakit HIV/AIDS sebesar 98.11765 kasus. Kasus penyakit IMS sebesar 114.2353. Kasus penyakit DBD sebesar 447.5294. Kasus penyakit Diare sebesar 83277.353. Kasus penyakit TB sebesar 770.4118 dan jumlah penyakit Malaria sebesar 8.11764. Dari hasil rata-rata objek yang terdapat dalam klaster 1 pada setiap variabel banyaknya kasus HIV/AIDS, IMS, DBD, TB, Malaria didefinisikan sedang dan banyaknya kasus Diare didefinisikan tinggi.

Pada klaster 2 rata-rata kasus penyakit HIV/AIDS sebesar 155.875 kasus. Kasus penyakit IMS sebesar 395. Kasus penyakit DBD sebesar 728.375. Kasus penyakit Diare sebesar 20464.88.

Kasus penyakit TB sebesar 1016.375 dan jumlah penyakit Malaria sebesar 13.25. Dari hasil rata-rata objek yang terdapat dalam klaster 2 pada setiap variabel banyaknya kasus HIV/AIDS, DBD, Diare, TB didefinisikan sedang dan banyaknya kasus IMS dan Malaria didefinisikan tinggi.

Pada klaster 3 rata-rata kasus penyakit HIV/AIDS sebesar 504.2 kasus. Kasus penyakit IMS sebesar 741.4. Kasus penyakit DBD sebesar 1195.2. Kasus penyakit Diare sebesar 60166.6. Kasus penyakit TB sebesar 3171.6 dan kasus penyakit Malaria sebesar 17.8. Dari hasil rata-rata objek yang terdapat dalam klaster 3 pada setiap variabel banyaknya kasus HIV/AIDS, IMS, DBD, Diare, TB, Malaria didefinisikan tinggi.

Pada klaster 4 rata-rata kasus penyakit HIV/AIDS sebesar 142.875 kasus. Kasus penyakit IMS sebesar 293.125. Kasus penyakit DBD sebesar 741. Kasus penyakit Diare sebesar 32391.5. Kasus penyakit TB sebesar 1299 dan jumlah penyakit Malaria sebesar 2.25. Dari hasil rata-rata objek yang terdapat dalam klaster 4 pada setiap variabel banyaknya kasus HIV/AIDS, DBD, TB, Malaria didefinisikan sedang dan banyaknya kasus IMS dan Diare didefinisikan tinggi.

2. Data 2

Tabel 4.35 Nilai Rata-Rata Objek Dalam Klaster Pada Setiap Variabel Data 2

Variabel	Klaster 1	Klaster 2	Klaster 3
Pertumbuhan Ekonomi	7.751	7.39	6.375
Kontribusi Terhadap Provinsi	3.018	33.52	1.535
IPM	65.67	79.35	75.65

Berdasarkan tabel 4.35 dapat dijabarkan bahwa pada klaster 1 terdiri dari Kabupaten/Kota di Sulawesi Selatan dengan rata-rata pertumbuhan ekonomi sebesar 7.751 persen. Rata-rata kontribusi terhadap provinsi sebesar 3.018 persen. Rata-rata Indeks Pembangunan Manusia sebesar 65.67 persen. Dari hasil rata-rata objek yang terdapat dalam klaster 1 pada setiap variabel indikator ekonomi pada pertumbuhan ekonomi, kontribusi terhadap provinsi dan IPM didefinisikan sedang.

Pada klaster 2 rata-rata pertumbuhan ekonomi sebesar 7.39 persen. Rata-rata kontribusi terhadap provinsi sebesar 33.52 persen. Rata-rata Indeks Pembangunan Manusia sebesar 79.35 persen. Dari hasil rata-rata objek yang terdapat dalam klaster 2 pada setiap variabel indikator ekonomi pada pertumbuhan ekonomi didefinisikan sedang dan kontribusi terhadap provinsi, IPM didefinisikan tinggi.

Pada klaster 3 rata-rata pertumbuhan ekonomi sebesar 6.375 persen. Rata-rata kontribusi terhadap provinsi sebesar 1.535 persen. Rata-rata Indeks Pembangunan Manusia sebesar 75.65 persen. Dari hasil rata-rata objek yang terdapat dalam klaster 3 pada setiap variabel indikator ekonomi pada pertumbuhan ekonomi, kontribusi terhadap provinsi didefinisikan rendah dan IPM didefinisikan tinggi.

3. Data 3

Tabel 4.36 Nilai Rata-Rata Objek Dalam Klaster Pada Setiap Variabel Data 3

Variabel	Klaster 1	Klaster 2
Tanaman Pangan	50.254	2.620
Hortikulturar	66.388	485.447
Perkebunan	58.129	12.004
Peternakan	134.982	586.667
Perikanan	66.577	622
Kehutanan	85.957	681.333

Berdasarkan tabel 4.36 dapat dijabarkan bahwa pada klaster 1 terdiri dari provinsi-provinsi di Indonesia dengan rata-rata subsektor tanaman pangan sebesar 50.254 pelaku usaha. Subsektor hortikultural sebesar 66.388 pelaku usaha. Subsektor perkebunan sebesar 58.129 pelaku usaha. Subsektor peternakan sebesar 134.982 pelaku usaha. Subsektor perikanan sebesar 66.577 pelaku usaha dan subsektor kehutanan sebesar 85.957 pelaku usaha. Dari hasil rata-rata objek yang terdapat dalam klaster 1 pada setiap variabel pelaku usaha pertanian pada subsektor peternakan, perikanan didefinisikan sedang dan subsektor tanaman pangan, hortikultural, perkebunan, kehutanan didefinisikan tinggi.

Pada klaster 2 rata-rata subsektor tanaman pangan sebesar 2.620 pelaku usaha. Subsektor hortikultural sebesar 485.447 pelaku usaha. Subsektor perkebunan sebesar 12.004 pelaku usaha. Subsektor

peternakan sebesar 586.667 pelaku usaha. Subsektor perikanan sebesar 622 pelaku usaha dan subsektor kehutanan sebesar 681.333 pelaku usaha. Dari hasil rata-rata objek yang terdapat dalam klaster 2 pada setiap variabel palaku usaha pertanian pada subsektor tanaman pangan didefinisikan rendah, subsektor perkebunan didefinisikan sedang dan subsektor peternakan, perikanan holtikultural, kehutanan didefinisikan tinggi.

4. Data 4

Tabel 4.37 Nilai Rata-Rata Objek Dalam Klaster Pada Setiap Variabel

Variabel	Klaster 1	Klaster 2
TK	15.103	31.24
SD	14.382	18.35
SMP	12.244	13.07
SMA	11.495	10.35
SMK	10.086	9.48
MI	12.157	11.84
MTS	11.657	9.96
MA	8.962	11.41

Berdasarkan tabel 4.37 dapat dijabarkan bahwa pada klaster 1 terdiri dari Kabupaten/Kota di Jawa Timur dengan rata-rata rasio murid dan guru TK sebesar 15.103. Rasio murid dan guru SD sebesar 14.382. Rasio murid dan guru SMP sebesar 12.244. Rasio murid dan guru SMA sebesar 11.495. Rasio murid dan guru SMK sebesar 10.086. Rasio murid dan guru MI sebesar 12.157. Rasio murid dan guru MTS sebesar 11.657. Rasio murid dan guru MA sebesar 8.962. Dari hasil rata-rata objek yang terdapat dalam klaster 1 pada setiap variabel banyaknya rasio murid dan guru TK, SD, SMP, SMA, SMK, MI, MTS, MA didefinisikan sedang.

Pada klaster 2 rata-rata rasio murid dan guru TK sebesar 31.24. Rasio murid dan guru SD sebesar 18.35. Rasio murid dan guru SMP sebesar 13.07. Rasio murid dan guru SMA sebesar 10.35. Rasio murid dan guru SMK sebesar 9.48. Rasio murid dan guru MI sebesar 11.84. Rasio murid dan guru MTS sebesar 9.96. Rasio murid dan guru MA sebesar 11.41. Dari hasil rata-rata objek yang terdapat dalam klaster 2 pada setiap variabel SMP, SMA, SMK, MI, MTS

didefinisikan sedang dan banyaknya rasio murid dan guru TK, SD, MA didefinisikan tinggi.



BAB V PENUTUP

5.1 Kesimpulan

Kesimpulan yang dapat diambil berdasarkan hasil analisis pada penelitian ini adalah sebagai berikut:

1. Hasil pengklasteran menggunakan indeks validitas *Gap Statistic* pada metode *average linkage*. Pada data 1 didapatkan banyaknya kluster adalah 4 kluster dan data 2 didapatkan banyaknya kluster adalah 3 kluster. Pada data 3 dan 4 didapatkan banyaknya kluster adalah 2. Hasil pengklasteran menggunakan indeks validitas *Silhouette* pada metode *average linkage*. Pada data 1, 2, 3 dan 4 didapatkan banyaknya kluster adalah 2 kluster. Hasil pengklasteran menggunakan indeks validitas *Davies-Bouldin* pada metode *average linkage*. Pada data 1 dan 2 didapatkan banyaknya kluster adalah 2 kluster dan pada data 3 dan 4 didapatkan banyaknya kluster adalah 5.
2. Hasil pengklasteran menggunakan metode *average linkage*, berdasarkan hasil nilai CTM terkecil menggunakan empat data menunjukkan bahwa indeks validitas *Gap Statistic* lebih baik dalam penentuan banyaknya kluster optimal dibandingkan dengan indeks validitas *Silhouette* dan *Davies-Bouldin* untuk keempat data tersebut.
3. Berdasarkan nilai indeks stabilitas dari keempat data menunjukkan bahwa rata-rata nilai indeks stabilitasnya dapat dikatakan stabil.

5.2 Saran

Saran dari penelitian ini adalah :

1. Dalam menentukan kluster optimal indeks validitas yang akan digunakan pada analisis kluster metode *average linkage* disarankan mempertimbangkan indeks validitas *gap statistic*.
2. Untuk penelitian selanjutnya disarankan menambah metode indeks validitas lain seperti Calinski-Harabasz, Hartigan, Dunn.



DAFTAR PUSTAKA

- Abdurrahan, M. 2016. Analisis *Cluster* untuk mengklasifikasikan Sektor Pertanian Berdasarkan Data Jumlah Rumah Tangga Usaha Pertanian Indonesia. *Skripsi*. Universitas Islam Indonesia. Yogyakarta.
- Badan Pusat Statistik. 2017. *Jawa Timur Dalam Angka*. BPS Provinsi Jawa Timur.
- Charrad, M., Ghazzali, N., Boiteau, V dan Niknafs, A., 2014. An R Package for Determining the Relevant Number of Clusters in a Data set. *Jurnal of Statistical Software*. 61,6.
- Everitt, B.S., Landau, S., Leese, M., dan Stahl, M. 2011. *Cluster Analysis*. Jhon Wiley & Sons, Ltd. United Kingdom.
- Epps, J dan Ambikairajah, E. 2008. Visualisation of Reduced-Dimension Microarray Data Using Gaussian Mixture Models. University of New South Wales, Sydney 2052. Australia.
- Fitrisiana. 2009. Perbandingan Metode Gap-Statistik Dan Indeks Validitas Gabungan (IVG) dalam Menentukan Banyaknya Cluster Optimal pada Analisis Cluster Menggunakan Data Ekspresi Gen. *Skripsi*. Universitas Brawijaya. Malang.
- Hair, J.F., Black, W.C., dan Anderson, R.E. 2010. *Multivariate Data Analysis Fifth Edition*. New Jersey: Prentice Hall Internasional, inc.
- Hennig, C. 2006. *Cluster-Wise Assessment of Cluster Stability*. University College London. United Kingdom.
- Hennig, C., Meila, M., Murtagh, F., dan Rocci, R. 2016. *Handbook of Cluster Analysis*. Taylor dan Francis Group, LLC.
- Jolliffe, I.T. 1986. *Principal Component Analysis*. Springer Verlag New York, inc.

- Johnson, R.A. dan Wichern, D.W. 2007. *Applied Multivariat Statistical Analysis*. New Jersey: Prentice Hall Internasional, inc.
- Kannan, K. S. dan Manoj, K. Outlier Detection In Multivariate Data. *Applied Mathematical Science*. Department of Statistics. India. 47, 2317-2324
- Rahayu, Sri dan Tarno. 2006. Prediksi Produksi Jagung di Jawa Tengah dengan ARIMA dan Bootstrap. *Jurnal Jurusan Matematika FMIPA UNDIP Semarang*. 979.704.427.0
- Sholihah, Fitriati. 2011. Penentuan Jumlah Kelompok Pada Analisis Kelompok Metode Single Linkage Menggunakan Beberapa Indeks Validitas. *Skripsi*. Universitas Brawijaya. Malang.
- Srinadi, I.G. A. M dan Sumarjaya I. W. 2016. Analisis Biplot untuk Pemetaan Posisi Karakteristik Usaha Pariwisata di Provinsi Bali. *Jurnal Matematika FMIPA UDAYANA*. 1693-1394
- Sukmawanti. 2017. Analisis *Cluster* dengan Metode Hierarki untuk Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Makro Ekonomi. *Skripsi*. Universitas Islam Negeri Alaudin. Makassar.
- Tibshirani, R., Walther, G., dan Hastie, T. 2001. Estimating The Number of Clusters in a Data Set Via The Gap Statistic. *J. R.Statist. Soc. B*. 411-423.